

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: February 27, 2005, 20:49:53 ; Search time 174 Seconds
(without alignments)
782.834 Million cell updates/sec
Title: US-10-805-311-24
Perfect score: 1401
Sequence: 1 MTAPVFSIIPTFNAAVTLQ.....ALRTRLIRKAVSKERSAEP 266

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 1612378 seqs, 512079187 residues
Total number of hits satisfying chosen parameters: 1612378
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : Uniprot_03: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1401	100.0	266	2	088109 mycobacteri
2	1401	100.0	266	2	07BR87 mycobacteri
3	1401	100.0	266	1	07B2P1 mycobacteri
4	788	56.2	256	1	YTS7 MYCTU
5	319	22.8	251	2	08A3K7 bacteroides
6	308	22.0	257	2	Q64WN8 bacteroides
7	298.5	21.3	297	2	Q93Q25 salmonella
8	298	21.3	297	2	Q9S520 escherichia
9	294	21.0	293	2	Q72WK9 desulfobibr
10	287	20.5	248	2	08P241 methanosarc
11	276	19.7	265	2	08VTX0 leptospira
12	275	19.6	265	2	Q8VW5 leptospira
13	275	19.6	265	2	08VX7 leptospira
14	275	19.6	265	2	Q9AEE4 leptospira
15	275	19.6	265	2	Q72Q14 leptospira
16	274	19.6	265	2	08VW9 leptospira
17	271	19.3	265	2	Q8VL41 leptospira
18	267.5	19.1	248	2	Q9EXY1 escherichia
19	267.5	19.1	258	2	Q7WYS0 rhizobium 1
20	267.5	19.1	265	2	08VW8 leptospira
21	266.5	19.0	248	2	Q9EXY4 escherichia
22	257.5	18.4	248	2	Q7MW87 porphyromon
23	255.5	18.2	251	2	Q9ZSK3 leptospira
24	255	18.2	254	2	Q9EXZ1 salmonella
25	255	18.2	274	2	Q7NEU1 gloeobacter
26	254.5	18.2	263	2	Q9S4F9 leptospira
27	252	18.0	254	2	Q9EXY9 salmonella
28	248	17.7	254	2	Q9XDJ3 bacteroides
29	247	17.6	276	2	Q9RQ94 rhizobium e
30	240	17.1	295	2	Q74BU3 geobacter s
31	238.5	17.0	247	2	Q66DN1 yersinia ps

32	238.5	17.0	247	2	Q9RCB0	Q9rcb0 yersinia ps
33	238.5	17.0	247	2	Q9RCC4	Q9rcC4 yersinia pe
34	238.5	17.0	341	2	Q6QW85	Q6qW85 azospirillum
35	235.5	16.8	257	2	Q64Q40	Q64q40 bacteroides
36	232.5	16.6	260	2	Q92MX1	Q92mx1 helicobacte
37	231.5	16.5	259	2	Q24928	Q24928 helicobacte
38	220	15.7	324	2	Q8YSL1	Q8ySl1 anabaena sp
39	207	14.8	321	2	Q8YSM2	Q8ySm2 anabaena sp
40	207	14.8	368	2	Q98JH2	Q98jH2 rhizobium 1
41	206.5	14.7	316	2	Q8YSL6	Q8ySl6 anabaena sp
42	205.5	14.7	298	2	Q9UZI6	Q9uZi6 pyrococcus
43	204.5	14.6	220	2	Q6QW98	Q6qW98 azospirillum
44	204.5	14.6	330	2	Q8YSM1	Q8ySm1 anabaena sp
45	203.5	14.5	255	2	Q7UVS5	Q7uV55 rhodopirell

ALIGNMENTS

RESULT 1
ID 088109 PRELIMINARY; PRT; 266 AA.
AC 088109;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DE 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Gsd protein.
GN Name=gad;
OS Mycobacterium avium subsp. silvaticum.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=44282;
RN [1]
RP SEQUENCE FROM N.A.
RA Tizard M., Bull T., Millar D., Doran T., Martin H., Ford J.,
RA Hermon-Taylor J.;
RT "A low G C content element in Mycobacterium avium subsp.
RT paratuberculosis and M. avium subsp. silvaticum with homologous genes
RT in M. tuberculosis";
RL Microbiology 144:3413-3423 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Bull T.;
RA Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ223832; CAA11574.1; -
DR InterPro: IPR001173; Glyco.trans.2.
DR Pfam; PF00535; Glycos.transf.2; 1.
SQ SEQUENCE 266 AA; 30195 MW; F88A2754693FSA8B CRC64;
Query Match 100.0%; Score 1401; DB 2; Length 266;
Best Local Similarity 100.0%; Pred. No. 1.3e-118;
Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MTAPVFSIIPTFNAAVTLQACLSIGVGTVEVWLVVDGSGTDTLTDIANSFPELGS	60
DB	1	MTAPVFSIIPTFNAAVTLQACLSIGVGTVEVWLVVDGSGTDTLTDIANSFPELGS	60
QY	61	RLVHSGPDGPDYDANNRGVATGEWLVFLGADDTLYEPTTLAQVAAFLGDDHAASHLVY	120
DB	61	RLVHSGPDGPDYDANNRGVATGEWLVFLGADDTLYEPTTLAQVAAFLGDDHAASHLVY	120
QY	121	GDVVMRSKSRHAGPFDLRLFLFETNLCHQSIFYRRELFDGIGPYNLRVYRWADDFNIR	180
DB	121	GDVVMRSKSRHAGPFDLRLFLFETNLCHQSIFYRRELFDGIGPYNLRVYRWADDFNIR	180
QY	181	CFSNPALITRYMDVVISYNDMTGFSMRQGTDKFRKRLPMYFWAGWETCRMLAFLKD	240
DB	181	CFSNPALITRYMDVVISYNDMTGFSMRQGTDKFRKRLPMYFWAGWETCRMLAFLKD	240
QY	241	KENRRLALRTRLIRKAVSKERSAEP 266	
DB	241	KENRRLALRTRLIRKAVSKERSAEP 266	

RESULT 2

Q7BR87 PRELIMINARY; PRT; 266 AA.
AC Q7BR87;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Glycosyltransferase gtfD.
GN Name=gtfD;
OS Mycobacterium avium.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1764;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=TMC 724, and 2151;
RX MEDLINE=22885473; PubMed=14523113; DOI=10.1099/mic.0.26528-0;
RA Eckstein T.M., Belisle J.T., Inamine J.M.;
RT "Proposed pathway for the biosynthesis of serovar-specific
glycopeptidolipids in Mycobacterium avium serovar 2.";
RL Microbiology 149:2797-2807(2003).
DR EMBL; AF125999; AAD20376.1; -;
DR EMBL; AF143772; AAD44222.1; -;
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR001173; Glyco_transf_2;
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Transferase.
SQ SEQUENCE 266 AA; 30195 MW; F88A2754683F5A8B CRC64;

Query Match 100.0%; Score 1401; DB 2; Length 266;
Best Local Similarity 100.0%; Pred. No. 1.3e-118;
Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTAPVFSIIIPTFNAAVTLQACLSIVGQTYREVEVVLVDGSGTDTLTLDIANSFRPELGS 60
DB 1 MTAPVFSIIIPTFNAAVTLQACLSIVGQTYREVEVVLVDGSGTDTLTLDIANSFRPELGS 60
QY 61 RLIVHSGPDDGPDYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAAPLGDHAAASHLVY 120
DB 61 RLIVHSGPDDGPDYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAAPLGDHAAASHLVY 120
QY 121 GDVWRSTKSRHAGPDLDRLLFETNLCHQSIFYRRELFDGIGPYNLRVYRWADWDNFIR 180
DB 121 GDVWRSTKSRHAGPDLDRLLFETNLCHQSIFYRRELFDGIGPYNLRVYRWADWDNFIR 180
QY 181 CFSNPALITRYMDVVISYNDMTGFSMRQGTDKFKRLPMYFWVAGWETCRMLAFLKD 240
DB 181 CFSNPALITRYMDVVISYNDMTGFSMRQGTDKFKRLPMYFWVAGWETCRMLAFLKD 240
QY 241 KENRRALRTRLIIRKAVSKERSAEP 266
DB 241 KENRRALRTRLIIRKAVSKERSAEP 266

RESULT 3

Q7B2P1 PRELIMINARY; PRT; 266 AA.
AC Q7B2P1;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Gsd protein (Hypothetical protein).
GN Name=gsd; OrderedLocusNames=MAP1234;
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1770;
[1]
RN SEQUENCE FROM N.A.
RC Tizard M., Bull T., Millar D., Doran T., Martin H., Ford J.,
RA Hermon-Taylor J.;
RT "A low G C content element in Mycobacterium avium subsp.

RT paratuberculosis and M. avium subsp. silvaticum with homologous genes
in M. tuberculosis.";
RL Microbiology 144:3413-3423(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Bull T.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=k10;
RA Li L., Bannantine J., Zhang Q., Amonsin A., Alt D., Kapur V.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ223833; CAA11578.1; -;
DR EMBL; AE017231; AAS03551.1; -;
DR InterPro; IPR001173; Glyco_transf_2;
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 266 AA; 30195 MW; F88A2754683F5A8B CRC64;

Query Match 100.0%; Score 1401; DB 2; Length 266;
Best Local Similarity 100.0%; Pred. No. 1.3e-118;
Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTAPVFSIIIPTFNAAVTLQACLSIVGQTYREVEVVLVDGSGTDTLTLDIANSFRPELGS 60
DB 1 MTAPVFSIIIPTFNAAVTLQACLSIVGQTYREVEVVLVDGSGTDTLTLDIANSFRPELGS 60
QY 61 RLIVHSGPDDGPDYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAAPLGDHAAASHLVY 120
DB 61 RLIVHSGPDDGPDYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAAPLGDHAAASHLVY 120
QY 121 GDVWRSTKSRHAGPDLDRLLFETNLCHQSIFYRRELFDGIGPYNLRVYRWADWDNFIR 180
DB 121 GDVWRSTKSRHAGPDLDRLLFETNLCHQSIFYRRELFDGIGPYNLRVYRWADWDNFIR 180
QY 181 CFSNPALITRYMDVVISYNDMTGFSMRQGTDKFKRLPMYFWVAGWETCRMLAFLKD 240
DB 181 CFSNPALITRYMDVVISYNDMTGFSMRQGTDKFKRLPMYFWVAGWETCRMLAFLKD 240
QY 241 KENRRALRTRLIIRKAVSKERSAEP 266
DB 241 KENRRALRTRLIIRKAVSKERSAEP 266

RESULT 4

YT57_MYCTU STANDARD; PRT; 256 AA.
ID YT57_MYCTU
AC Q50459;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Putative glycosyl transferase Rv2957/MT3031/Mb2981 (EC 2.-.-.-).
GN OrderedLocusNames=Rv2957, MT3031, Mb2981;
ORFNames=MTCY349.31c, u0002kc;
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
[1]
RN SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis;
RA Smith D.R., Robison K.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230; DOI=10.1038/31159;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C.M.,
RA Harris D.E., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III,
RA Tekala F., Badcock K., Basham D., Brown D., Chillingworth T.,
RA Connor R., Davies R.M., Devlin K., Feltwell T., Gentles S., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S.,

```
RA Murphy L.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,
RA Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulten J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence."
RL Nature 393:537-544(1998).
[3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
EX DOI=10.1128/JB.184.19.5479-5490.2002;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.D., DeBoy R.T., Dodson R.J., Gwinn M.L., Haft D.H.,
RA Hickey E.K., Kolonay J.F., Nelson W.C., Umayama M.D.,
RA Salzberg S.L., Delcher A., Utterback T.R., Weidman J.P., Khouri H.M.,
RA Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C.,
RA Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
[4]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;
RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Gironde S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -!- SIMILARITY: Belongs to the glycosyltransferase 2 family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U00024; AA05038.1; ALT_INIT.
DR EMBL; BX842581; CAB05419.1; ALT_INIT.
DR EMBL; AE000516; AAK47357.1; -.
DR EMBL; BX248344; CAD96668.1; ALT_INIT.
DR TIGR; MT3031; -.
DR TubercuList; Rv2957; -.
DR InterPro; IPR001173; Glyco_trans_2.
DR Pfam; PF00535; Glycosyltransferase_2; 1.
KW Complete proteome; Glycosyltransferase; Hypothetical protein;
KW Transferase.
SQ SEQUENCE 256 AA; 29012 MW; 90C6EC628C59CA57 CRC64;

Query Match 56.2%; Score 788; DB 1; Length 256;
Best Local Similarity 60.9%; Pred. No. 4.4e-63;
Matches 156; Conservative 24; Mismatches 54; Indels 22; Gaps 2;

QY 1 MTPVFSIIIPTFNAAVTLOACLGSIQVGYREYEVVLVDGSGTDRTLDIANSFPELGS 60
DB 1 MAAPNFSIIIPILNVAVLPACLDISARQTCGDFELVLDGSGTDRTLDIANSFPELGS 60

QY 61 RLIVHSGPDDGPDYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAFLGDHAAHLVY 120
DB 61 RLIIHRDTDQGYDAMNRGVLDATGTLWLLFGADDSLYEADTLARVAFIGEHEPSDLVY 120

QY 121 GDVVRSTKSRHAGPFDRLLLFETNLCHQSIFYRRELFDGIGPNLYRVRWADFNIR 180
DB 121 GDVIMRSTNFRWGAFLDRLFLKKNICHOAIFYRRGLFGTIGPNLYRVRWADFNIR 180

QY 181 CFSNPALITRYMDVVIYSEYND--MTGFSMRQGTDKERKRLPMYFWA-----G 227
DB 181 CFSNPALITRYMHHVVASNEFGGLS-NTIVDKERKRLPM----- 220

QY 241 KENRRALRLTRILRVK 256
```

Db 221 --STRLGIRLVIVLVR 234

RESULT 5

```
Q8A3K7 PRELIMINARY; PRT; 251 AA.
AC Q8A3K7
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative glycosyltransferase.
GN OrderedLocustNames=BT2947;
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12663928; DOI=10.1126/science.1080029;
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis."
RL Science 299:2074-2076(2003).
DR EMBL; AE016938; AAO78053.1; -.
DR HSSP; P39621; 10QO.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR001173; Glyco_trans_2.
DR Pfam; PF00535; Glycosyltransferase_2; 1.
KW Complete proteome; Transferase.
SQ SEQUENCE 251 AA; 28977 MW; 78C7BFBC9C13F067 CRC64;
```

Query Match 22.8%; Score 319; DB 2; Length 251;

Best Local Similarity 33.1%; Pred. No. 1.4e-20;

Matches 82; Conservative 41; Mismatches 101; Indels 24; Gaps 8;

QY 4 PVFSIIIPTFNAAVTLOACLGSIQVGYREYEVVLVDGSGTDRTLDIANSFPELGSRLV 63

DB 8 PKFSIIITVYNAEKVLEDTIQSVISQTYHHVEYIIVDGASKDGTLSIIDRYPRI---TT 64

QY 64 VHSOPDDGPDYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAFLGDHAAHLVYGDV 123

DB 65 VVSEPDKGLYDAMNKAISLASGDYLCFLNAGDCFEDDTLQOMVHSHGSLVLDVIYGET 124

QY 124 VMRSTKSRH-----AGPPDLRLPETNL--CHQSIFYRRELFDGIGPNLYRVRWAD 174

DB 125 AI-VDKDRHFLHMRRLSAPEKLWKSFGQGLVCHQAFPAHRTL---VEPIDLSYRYGAD 180

QY 175 WDFNIRCFSPALITRYMDVVIYSEYND--MTGFSMRQGTDKERKRLPMYFWA-----G 227

DB 181 FDCIRIMKK-AYTLHNTHLTIIIDYLDGEMTTQNRKASLKERFRIMAKHYGWIATAAHA 239

QY 228 WETCRML 235

DB 240 WFLRLII 247

RESULT 6

```
Q64MN8 PRELIMINARY; PRT; 257 AA.
AC Q64MN8
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Putative glycosyltransferase.
GN ORFNames=BF4512;
OS Bacteroides fragilis.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=817;
RN [1]
RP SEQUENCE FROM N.A.
```


[illegible]

Job time : 177 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2005, 21:01:28 ; Search time 40 Seconds
(without alignments)
639.841 Million cell updates/sec

Title: US-10-805-311-24
Perfect score: 1401
Sequence: 1 MTAPVFSIIPTFNAATLQ.....ALRTLIRKAVSKERSAEP 266
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	788	56.2	275	2 B70670	hypothetical prote
2	238.5	17.0	247	2 AG0376	probable glycosylt
3	232.5	16.6	260	2 E71975	hypothetical prote
4	231.5	16.5	259	2 F64532	conserved hypothet
5	220	15.7	324	2 AB2190	hypothetical prote
6	207	14.8	321	2 AG2188	hypothetical prote
7	206.5	14.7	316	2 AE2189	hypothetical prote
8	205.5	14.7	298	2 B75096	glycosyl transfera
9	204.5	14.6	330	2 AH2188	hypothetical prote
10	202.5	14.5	262	2 E70714	hypothetical prote
11	196.5	14.0	337	2 AG1920	hypothetical prote
12	196.5	14.0	367	2 G95948	probable glycosylt
13	192	13.7	336	2 AE7168	glycosyltransferas
14	190.5	13.6	322	2 AE2160	glycosyltransferas
15	190.5	13.6	323	2 AD2189	hypothetical prote
16	190	13.6	248	2 C90984	probable glycosylt
17	190	13.6	248	2 F85829	glycosyl transfera
18	189.5	13.5	333	2 AH2026	hypothetical prote
19	188.5	13.5	333	2 B97168	glycosyltransferas
20	188	13.4	299	2 B83557	probable glycosylt
21	182	13.0	344	2 AC0974	probable glycosylt
22	178	12.7	356	2 S74766	hypothetical prote
23	177.5	12.5	343	2 AI2091	glycosyltransferas
24	175	12.5	318	2 AG2189	hypothetical prote
25	174.5	12.5	333	2 G86651	sugar transferase
26	174.5	12.5	623	2 AH1209	B. subtilis minor
27	174.5	12.5	996	2 S76194	hypothetical prote
28	173.5	12.4	324	2 AE9290	probable glycosylt
29	173.5	12.4	338	2 E91190	probable regulator

ALIGNMENTS

RESULT 1

B70670	Hypothetical protein Rv2957 - Mycobacterium tuberculosis (strain H37RV)
N:Alternate names: u0002kc protein	
C:Species: Mycobacterium tuberculosis	
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000	
C:Accession: B70670; S73064	
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, A.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998	
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.	
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome	
A:Reference number: A70500; MUID:98295987; PMID:9634230	
A:Accession: B70670	
A>Status: preliminary; nucleic acid sequence not shown; translation not shown	
A:Molecule type: DNA	
A:Residues: 1-275 <COL>	
A:Cross-references: GB:Z83018; GB:AL123456; NID:g3261671; PIDN:CAB05419.1; PID:g1694875	
A:Experimental source: strain H37RV	
R:Smith, D.R.; Robison, K.	
A:Submitted to the EMBL Data Library, September 1994	
A:Description: Mycobacterium tuberculosis cosmid tbc2.	
A:Reference number: S73053	
A:Accession: S73064	
A>Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-275 <SMI>	
A:Cross-references: EMBL:U00024; NID:g560506; PIDN:AAA50938.1; PID:g560517	
C:Genetics:	
A:Gene: Rv2957	
C:Superfamily: Neisseria meningitidis glycosyl transferase A	
Query Match 56.2%; Score 788; DB 2; Length 275;	
Best Local Similarity 60.9%; Pred. No. 4.1e-65;	
Matches 156; Conservative 24; Mismatches 54; Indels 22; Gaps 2;	
QY 1 MTAPVFSIIPTFNAATLQACLSIVCQTVREVEVLVDGSGTDTLDIANSFPELGS 60	probable regulator
DB 20 MAAPVFSIIPTLNVAALVPACLDSIARQTCGDFELVLVDGSGTDTLDIANIFAPNLGE 79	hypothetical 40.5k
QY 61 RLNVHSGPDDGYDAMNNGVGVATGEWLFGLGADTLVEPTTLAQVAFLGHPAASHLVY 120	hypothetical prote
DB 80 RLIIHRDTQGVYDMNNGVDLATGTWLLFGLGADSLYEADTLARVAALFIGHESDLVY 139	capsular polysacch
QY 121 GDVVMRSTKSRHAGPFDLRLFFETNLCHQISFYRRELFDDGIGPYNLRYRWADDFNIR 180	glycosyl transfera
DB 140 GDVIMRSTNFRWGGAFLDRLLFKRNICHQALFYRGLFGTIGPYNLRYRLVADDFNIR 199	probable glycosylt
QY 181 CFSNPALITRMVDVVISEYNDMTGSMRQGTDKPRKRLPMYFVWAGWETCRMLAFLKD 240	probable glycosyl
DB 200 CFSNPALVTRYMHVVVASYNEFGGLS-NTIIVDKFELKRLPM----- 239	hypothetical prote

QY 241 KENRRLALRTRLIRVK 256
Db 240 --STRGLRVLVLVR 253

RESULT 2

AG0376
probable glycosyltransferase wbyL [imported] - Yersinia pestis (strain C092)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C:Accession: AG0376
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titchell, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
ii, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AG0376
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-247 <KUR>
A:Cross-references: UNIPROT:Q9RCC4; GB:AL590842; PIDN:CAC92338.1; PID:g15981048; GSPDB:Q
C:Genetics:
A:Gene: wbyL

Query Match 17.0%; Score 238.5; DB 2; Length 247;
Best Local Similarity 32.2%; Pred. No. 28-14;
Matches 65; Conservative 41; Mismatches 75; Indels 21; Gaps 6;

QY 7 SIIIPTFNAAVTLQACLSIGVQTYREVWVLVDGSGDTRTLDI--ANSFRPELGSRLVW 64
Db 4 SIIITATYNGERTISDTLSLEKQTVLDVEYIIVDGAASKNTLTVISQNSTRV---TKIIS 60

QY 65 HSGPDGPGYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAFLGDHAASHLVYGDV- 123
Db 61 ES--DQGIYDALNKGIDLATGDIIGFLHSDDLLAYPGALSDIVETF-EKQQCDAVYGDQ 117

QY 124 -----VMRSTKSRHAGPFDRLRLLFETNLCHQSIFYRRELFDGIGIPYRLRYRWADWDFNI 174
Db 118 YVAKNDLNNKVRWCWS---GFFNHKMKYGMWPPHPTFYMKRDLYIRFGGFDLSYKISAD 174

QY 175 WDFNRCFSNPALITRYMDWVI 196
Db 175 YDSLTRYIINYKIAYLPKVI 196

RESULT 3

E71975
hypothetical protein jhp0094 - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C:Accession: E71975
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: E71975
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-260 <ARN>
A:Cross-references: UNIPROT:Q9ZMK1; GB:AE001448; GB:AE001439; NID:g4154594; PIDN:AAD0567
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0094

Query Match 16.6%; Score 232.5; DB 2; Length 260;
Best Local Similarity 25.0%; Pred. No. 7.5e-14;
Matches 62; Conservative 53; Mismatches 94; Indels 39; Gaps 6;

QY 7 SIIIPTFNAAVTLQACLSIGVQTYREVWVLVDGSGDTRTLDIANSFRPELGSRLVWHS 66

Db 5 SVITACFNSEKTIETDILSVLNQYKNIEYIIIDGASADSTLEIIQKYKDRIA---CVMS 61
QY 67 GPDDGPGYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAFLGDHAASHLVYGDV--- 123
Db 62 EKDEGIYDAMNKGIRSSGDIIALNSDD-FYKDFVVEKVVHFENKNCDSVYDGLGV 120

QY 124 ---VMRSTKSRHAGPFDRLRLLFETNLCHQSIFYRRELFDGIGIPYRLRYRWADWDFNI 179
Db 121 KPDCLKVVRYVEGFSFKTLLYGVPVPAHPTLFVKKAIERYGLYKTDYKISADFEMI 180

QY 180 RCFNSNPALITRYMDVWVISEYNDMTGFSMRQG--TDKEFRKRLPMYFWVAGWETCRMLAF 237
Db 181 RLFFVQKISFSLKVEL-----VINRTGGVSASGFKSL----- 214

QY 238 LKDKENRR 245
Db 215 LRNKENLR 222

RESULT 4

F64532
conserved hypothetical protein HP0102 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
C:Accession: F64532
R:Tombs, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khairak, H.G.; Glodek, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: F64532
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-259 <TOM>
A:Cross-references: UNIPROT:O24928; GB:AE000532; GB:AE000511; NID:g2313184; PIDN:AAD0717
C:Genetics:
A:Start codon: TTG

Query Match 16.5%; Score 231.5; DB 2; Length 259;
Best Local Similarity 27.4%; Pred. No. 9.2e-14;
Matches 54; Conservative 46; Mismatches 86; Indels 11; Gaps 3;

QY 7 SIIIPTFNAAVTLQACLSIGVQTYREVWVLVDGSGDTRTLDIANSFRPELGSRLVWHS 66
Db 5 SVITACFNSEKTIETDILSVLNQYKNIEYIIIDGASDSTLEIIQKRDRIA---CVMS 61

QY 67 GPDDGPGYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAFLGDHAASHLVYGDVW- 125
Db 62 EKDEGIYDAMNKGIRSSGDIIALNSDD-FYKDFVVEKVVHFENKNCDSVYADLVV 120

QY 126 -----RSTKSRHAGPFDRLRLLFETNLCHQSIFYRRELFDGIGIPYRLRYRWADWDFNI 179
Db 121 KPDCLKVVRYVEIGEFPKTLTYGVVPAHPTLFVKKAIERYGLYKTDYKISADFEMI 180

QY 180 RCFNSNPALITRYMDWVI 196
Db 181 RLFFVQKISFSLKVEL 197

RESULT 5

AB2190
hypothetical protein alr3073 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AB2190
R:Kaneke, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi
Nakazaki, N.; Shampo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AB2190

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-324 <KUR>

A;Cross-references: UNIPROT:Q8YSL1; GB:BA000019; PIDN:BA74772.1; PID:gl7132167; GSPDB:

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: alr3073

C;Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match 15.7%; Score 220; DB 2; Length 324;

Best Local Similarity 25.9%; Pred. No. 1.4e-12;

Matches 69; Conservative 53; Mismatches 108; Indels 36; Gaps 8;

QY 4 PVFSIIPTFNAAVTLQACLGSIQVGTREVEVLVDGSGTDTTLDIANSFRPELGSRLV 63

DB 2 PKISVILPAYNAERTILETINSVLNQTFSDELIIVINDGSTDTREVLQVND---DARLK 58

QY 64 VHSPPDGPYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAFLGDHAAASHLYVGDV 123

DB 59 VYSYNSRASCARNHGISHAVGDFISFLDADD-LWTFDKLEQLSALNNHPEAGVAYSWT 117

QY 124 VNRSTKSRHAGPFD-----LDRL--FETNLCHQSIFYRRELFDGIGPNLRYRW 173

DB 118 YTIDDKGELLKPEPLVEGNYVTDLLANFLTNGSNPLI--RKAIAISIGFDTLRSGE 175

QY 174 DWDFNIRCFSPNALITRYMDVWVISEYNDMTGFSMRQGTDEKFRKRLPMYFWAGWETCRR 233

DB 176 DWDYWLRL-----LAKYFVFWVQKHQIL---YRRSVTSKSPKLQI-----IRE 215

QY 234 MIAFLDKENRRRLALTRILIRKAVS 259

DB 216 ASLAILDKAMKVLPLELQYLKXHSLS 241

RESULT 6

AG2188

hypothetical protein alr3062 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004

C;Accession: AG2188

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AG2188

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-321 <KUR>

A;Cross-references: UNIPROT:Q8YSM2; GB:BA000019; PIDN:BA74761.1; PID:gl7132156; GSPDB:

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: alr3062

C;Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match 14.8%; Score 207; DB 2; Length 321;

Best Local Similarity 32.0%; Pred. No. 2.2e-11;

Matches 66; Conservative 35; Mismatches 83; Indels 22; Gaps 7;

QY 4 PVFSIIPTFNAAVTLQACLGSIQVGTREVEVLVDGSGTDTTLDIANSFRPELGSRLV 63

DB 2 PKVSUVIPAYNAMPYLPETLESVLRQTYHDFVWVNDGSSDNTTEWWSQI---LDPLRK 58

QY 64 VHSPPDGPYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAFLGDHAAASHLYVGDV 123

DB 59 LISQANOGLAGARTNTGIVNASGEYIAFLDADD-IWEPTKLAQVSLDENPTVGLVYTWV 117

QY 124 VNRSTKSRHAGPFDLRL-----LFETNL--CHQSIFYRRELFDGIGPY--NLRV 172

DB 118 AYIDEQKSTGKIFKNQVEGYWMPQLTEHNIVECGSVALVRRVCPEKMGFLFORNLGSYY- 176

QY 173 ADWDFNIRCFSPNALITRYMDVWVISE 198

DB 177 EDMDWMLR-----IATSYDFKVKKE 196

RESULT 7

AE2189

hypothetical protein alr3068 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004

C;Accession: AE2189

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AE2189

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-316 <KUR>

A;Cross-references: UNIPROT:Q8YSL6; GB:BA000019; PIDN:BA74767.1; PID:gl7132162; GSPDB:

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: alr3068

C;Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match 14.7%; Score 206.5; DB 2; Length 316;

Best Local Similarity 29.6%; Pred. No. 2.4e-11;

Matches 55; Conservative 37; Mismatches 81; Indels 13; Gaps 5;

QY 4 PVFSIIPTFNAAVTLQACLGSIQVGTREVEVLVDGSGTDTTLDIANSFRPELGSRLV 63

DB 2 PTISVILPAYNAERTILETINSVQQQTFSDFLIINDGSTDTRELTQLNTRDE---RLK 58

QY 64 VHSPPDGPYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAFLGDHAAASHLYVGDV 123

DB 59 IFSYENGGLCTARNRGISHASGEFTAFLDADDLWTHDKLEQLTA-LQHPPEAGVAYSWT 117

QY 124 VNRSTKSRHAGP-----FDLD---RLLETNLCHQS--IFYRRELFDGIGPNLRYRW 174

DB 118 YFMDQSGKSSIPGVSLFFEGDVQAHLVNNFLASGNSPLIRKQAIYESVGFEFSDNCMGCAD 177

QY 175 WDFNIR 180

DB 178 WDYWLRL 183

RESULT 8

B75096

Glycosyl transferase PAB0772 - Pyrococcus abyssi (strain Orsay)

C;Species: Pyrococcus abyssi

C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004

C;Accession: B75096

R;anonymous, Genoscope

A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str

A;Reference number: A75001

A;Accession: B75096

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-298 <KAW>

A;Cross-references: UNIPROT:Q9UZI6; GB:AJ248286; GB:AL096836; NID:gs458366; PIDN:CAB500

A;Experimental source: strain Orsay

C;Genetics:

A;Gene: PAB0772

C;Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match 14.7%; Score 205.5; DB 2; Length 298;

Best Local Similarity 29.0%; Pred. No. 2.8e-11;

Matches 71; Conservative 38; Mismatches 99; Indels 37; Gaps 10;

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2005, 21:10:59 ; Search time 134 Seconds
(without alignments)
651.193 Million cell updates/sec

Title: US-10-805-311-24

Perfect score: 1401

Sequence: 1 MTAPVFSIIIFTFNAATLQ.....ALRTRLIRKVAKERSAEP 266

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA.*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09D_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	788	56.2	256	15 US-10-282-122A-62545	Sequence 62545, A
2	788	56.2	275	15 US-10-282-122A-64808	Sequence 64808, A
3	238.5	17.0	247	15 US-10-282-122A-78093	Sequence 78093, A
4	192	13.7	336	15 US-10-282-122A-51835	Sequence 51835, A
5	189.5	13.5	331	15 US-10-282-122A-51835	Sequence 51835, A
6	188.5	13.5	333	15 US-10-282-122A-51835	Sequence 51835, A
7	188	13.4	299	15 US-10-282-122A-66207	Sequence 66207, A
8	184.5	13.2	332	14 US-10-182-960-13	Sequence 13, Appl
9	182	13.0	291	15 US-10-282-122A-59156	Sequence 59156, A
10	182	13.0	327	15 US-10-282-122A-69114	Sequence 69114, A
11	182	13.0	344	15 US-10-282-122A-75381	Sequence 75381, A
12	180	12.8	343	15 US-10-282-122A-72785	Sequence 72785, A
13	179	12.8	278	9 US-09-767-041-36	Sequence 36, Appl

14	179	12.8	477	15	US-10-282-122A-67095	Sequence 67095, A
15	178	12.7	332	9	US-09-767-041-22	Sequence 22, Appl
16	177	12.6	421	15	US-10-282-122A-48233	Sequence 48233, A
17	176	12.6	360	15	US-10-282-122A-48802	Sequence 48802, A
18	174.5	12.5	623	15	US-10-282-122A-60382	Sequence 60382, A
19	173.5	12.4	331	15	US-10-282-122A-56382	Sequence 56382, A
20	172.5	12.3	344	15	US-10-282-122A-56744	Sequence 56744, A
21	172	12.3	358	9	US-09-815-242-5714	Sequence 5714, A
22	172	12.3	573	9	US-09-815-242-5714	Sequence 5714, A
23	172	12.3	573	15	US-10-282-122A-43816	Sequence 43816, A
24	172	12.3	573	17	US-10-857-625-785	Sequence 785, App
25	170.5	12.2	301	16	US-10-474-776-711	Sequence 711, App
26	170.5	12.2	301	17	US-10-472-928-3642	Sequence 3642, App
27	170.5	12.2	346	15	US-10-282-122A-57998	Sequence 57998, A
28	168	12.0	250	15	US-10-282-122A-58334	Sequence 58334, A
29	167.5	12.0	320	15	US-10-461-990-28	Sequence 28, Appl
30	167.5	12.0	1504	14	US-10-156-761-9772	Sequence 9772, Ap
31	165.5	11.8	150	9	US-09-924-358-29	Sequence 29, Appl
32	165.5	11.8	150	14	US-10-133-709-4	Sequence 4, Appli
33	165.5	11.8	150	15	US-10-410-764-29	Sequence 29, Appl
34	165.5	11.8	732	14	US-10-156-761-12627	Sequence 12627, A
35	165	11.8	251	9	US-09-738-626-4202	Sequence 4202, Ap
36	164	11.7	613	15	US-10-282-122A-50043	Sequence 50043, A
37	163.5	11.7	333	15	US-10-282-122A-52072	Sequence 52072, A
38	163.5	11.7	713	15	US-10-282-122A-42526	Sequence 42526, A
39	163	11.6	332	9	US-09-767-041-21	Sequence 21, Appl
40	162.5	11.6	274	9	US-09-738-626-3905	Sequence 3905, Ap
41	162.5	11.6	438	15	US-10-282-122A-60207	Sequence 60207, A
42	161.5	11.5	346	15	US-10-275-026A-26	Sequence 26, Appl
43	161	11.5	712	15	US-10-282-122A-57674	Sequence 57674, A
44	159.5	11.4	120	9	US-09-767-041-51	Sequence 51, Appl
45	159.5	11.4	696	17	US-10-472-928-3656	Sequence 3656, Ap

ALIGNMENTS

RESULT 1

US-10-282-122A-62545
; Sequence 62545, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931

```
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62545
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Mycobacterium bovis
US-10-282-122A-62545

Query Match      56.2%; Score 788; DB 15; Length 256;
Best Local Similarity 60.9%; Pred. No. 4e-79;
Matches 156; Conservative 24; Mismatches 54; Indels 22; Gaps 2;

QY  1  MTAPVFSIIPTFNAAVTLQACLSIGVGTREVEVVLVDGSGTDRTLDIANSRPELGS 60
Db  1  MAAPMFSIIPTLVNAAVLPACLDSIARQTCGDFELVLVDGSGTDETLDIANIFAPNLGE 60

QY  61  RLIVHSGPDDGPDYDAMNRGVGVATGEWVFLGADDTLYEPTTLAQVAFLGDHAAASHLVY 120
Db  61  RLIIHRDTDQGVYDAMNRGVDLATGCTWLLFLGADDSLYEADTLARVAAFIGEHEPSDLVY 120

QY  121  GDVVMRSKSHAGFPDRLLLFETNLCHQSIFYRRELFDDGIGPYNLRYRWADWDFNIR 180
Db  121  GDVIMRSTNFRWGAFLDRLLLFKRNICHQAIKYRRLGFGTIGTPYNLRYRVLDWDFNIR 180

QY  181  CFSNPALITRYMDVVISEYNDMTGFSMRQGTDKERKELPMYFWVAGWETCRRMLAFKLD 240
Db  181  CFSNPALVTRYMHVVVASYNEFGGLS-NTIIVDKFLKRLPM----- 220

QY  241  KENRELALRTRLIRVK 256
Db  221  --STRLGIRLVILVR 234

RESULT 2
US-10-282-122A-64808
; Sequence 64808, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Olsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
```

```
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64808
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-64808

Query Match      56.2%; Score 788; DB 15; Length 275;
Best Local Similarity 60.9%; Pred. No. 4.4e-79;
Matches 156; Conservative 24; Mismatches 54; Indels 22; Gaps 2;

QY  1  MTAPVFSIIPTFNAAVTLQACLSIGVGTREVEVVLVDGSGTDRTLDIANSRPELGS 60
Db  20  MAAPMFSIIPTLVNAAVLPACLDSIARQTCGDFELVLVDGSGTDETLDIANIFAPNLGE 79

QY  61  RLIVHSGPDDGPDYDAMNRGVGVATGEWVFLGADDTLYEPTTLAQVAFLGDHAAASHLVY 120
Db  80  RLIIHRDTDQGVYDAMNRGVDLATGCTWLLFLGADDSLYEADTLARVAAFIGEHEPSDLVY 139

QY  121  GDVVMRSKSHAGFPDRLLLFETNLCHQSIFYRRELFDDGIGPYNLRYRWADWDFNIR 180
Db  140  GDVIMRSTNFRWGAFLDRLLLFKRNICHQAIKYRRLGFGTIGTPYNLRYRVLDWDFNIR 199

QY  181  CFSNPALITRYMDVVISEYNDMTGFSMRQGTDKERKELPMYFWVAGWETCRRMLAFKLD 240
Db  200  CFSNPALVTRYMHVVVASYNEFGGLS-NTIIVDKFLKRLPM----- 239

QY  241  KENRELALRTRLIRVK 256
Db  240  --STRLGIRLVILVR 253

RESULT 3
US-10-282-122A-78093
; Sequence 78093, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Olsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
```



```
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78093
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Yersinia pestis
US-10-282-122A-78093

Query Match 17.0%; Score 238.5; DB 15; Length 247;
Best Local Similarity 32.2%; Pred. No. 8.3e-18;
Matches 65; Conservative 41; Mismatches 75; Indels 21; Gaps 6;

Qy 7 STIIPTFNAAVTLOACLSIGVQTYREVEVVLVDGSGTDRITDI--ANSFRPELGSRLV 64
Db 4 STIATYNSERTISTLSLEKQTYLDVEYIIVDGSNDTLTVISQSTRV---TKLIS 60
Qy 65 HSGPDGPDYDAMNRGVGVATGEWVLFGLGADDTLYEPTTLAQVAAPLGDHAAASHLVYGDV- 123
Db 61 ES--DOGIYDALNKIGIDLATGDIIGFLSHDOLLAYPGALS DIVERF-EKQCCDAYYGDQ 117
Qy 124 -----VMKSTKSRHAGPDLRLLFETNLCQSTFYRRELPGDGPYNLRVYRWAD 174
Db 118 YVAKNDLANKVRCWKS---GPFNHEKMKYGMWPPHPTFYMKRDLYIRFGGDLKYKISAD 174
Qy 175 WDFNRCFSPNALITRYMDVVI 196
Db 175 YDSLTRYIINIKIAYLPKVI 196

RESULT 4
US-10-282-122A-51835
; Sequence 51835, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51835
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-51835

Query Match 13.7%; Score 192; DB 15; Length 336;
Best Local Similarity 23.6%; Pred. No. 2e-12;
Matches 65; Conservative 57; Mismatches 113; Indels 40; Gaps 10;

Qy 4 PVFSIIPTFNAAVTLOACLSIGVQTYREVEVVLVDGSGTDRITDIANSFRPELGSRLV 63
Db 5 PKVISVMPVNSERYLAELIESILDQTYNDFEPIVDGSDTDESNNIISYANK-DNRII 63
Qy 64 VHSPPDDGPDYDAMNRGVGVATGEWVLFGLGADD-----TLYEPTTLAQVAAPLGDHAA 115
Db 64 VISREHRLVDLSNEGNIARAGKYIARMADADDISINNRIEKQFLELNKQVDIILGTRIE 123
Qy 116 SHLYVGVVWVST---KSRHAGPDLRL--LFTN--LCHQSIFYRRELFDGICPYNLR 168
Db 124 A---FGDIDEKQKTYNSAFSIFKFDSONIEQVFLTSCAIPHSVMFKKDSIVKLGGRKE 180
Qy 169 YRVWADWDENIRCFSPNALITRYMDVWVISEYN-----DMTGFMSRQG---- 210
Db 181 YDTAEDYDLWLRATENGKIVR-MDECLIKYRVNKSKTAVEMFNPKWVEYTMKAKIDYI 239
Qy 211 TDKEFRKRLPMYFWA--GWETCRRMLAFLKDKEN 243
Db 240 NDTNKKDKVDYLINGASTGGKLVKKVVESTTDKEN 274

RESULT 5
US-10-264-213-152
; Sequence 152, Application US/10264213
; Publication No. US20040009490A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka J
; APPLICANT: Lubbers, Mark William
; APPLICANT: Dekker, James
; TITLE OF INVENTION: Polynucleotides, materials incorporating
; FILE REFERENCE: 11000.1043C3
; CURRENT APPLICATION NUMBER: US/10/264,213
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 152
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Lactobacillus rhamnosus
US-10-264-213-152

Query Match 13.5%; Score 189.5; DB 15; Length 331;
Best Local Similarity 28.2%; Pred. No. 3.7e-12;
Matches 62; Conservative 38; Mismatches 93; Indels 27; Gaps 9;

Qy 4 PVFSIIPTFNAAVTLOACLSIGVQTYREVEVVLVDGSGTDRITDIANSFRPELGSRLV 63
Db 7 PLVSIIVPIYNNVEKYLQRCIDSLFAQTYVNIIEVLVDGSDRSLSICK-NAQODERV 65
Qy 64 VHSPPDDGPDYDAMNRGVGVATGEWVLFGLGADDTLYEPTTLAQVAAPL--GDHAAASHLVYGD 122
```

Db 66 VFSTKNGGVADTRNFGVSGVARQWISFVDPDDYV-DPDVIEYLGILVKSNAATSGICQHR 124
QY 123 VVMRSTK---SRHAGPFDLRLPETNLCHQSI-FYRRELFDGIGPNLYRWAD----W 175
Db 125 NVYKNGKIQTNLVEGPAVLDD-----SHTAV--KRLLYDD-----QIDTSVWAKLYPAW 170
QY 176 DNIRCFNSPALITRYMDVVISYNDMTGFSMRQGTDKF 215
Db 171 VFKKJHFPKGRLL---FEDIAATYKTFLASDSIAVGSEAKY.207

RESULT 6
US-10-282-122A-52169
; Sequence 52169, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-08-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52169
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-52169

Query Match 13.5%; Score 188.5; DB 15; Length 333;
Best Local Similarity 25.8%; Pred. No. 4.8e-12;
Matches 59; Conservative 42; Mismatches 95; Indels 33; Gaps 7;
QY 5 VFSIIPTFNAAVTLQACLSIGSVGTQYREVEVLVDGGSTDRTLTDLTANSFRPELGSRLV 64
Db 1 MISVMPVYCKEYLEESTESILKQTYRDFEIIIVDGSNDKSIDIIINKYANDNDNRIVV 60
QY 65 HSGPDGPDYDAMNRGVGATGWLFLGADDTLYEPTTLAQVAAPFLGDH-----AASHL 118
Db 61 SRDNNMGVYSLNIGIDRAKGSYARMDDADD-IALPERFERQIEYLNKNKQVDILACKVE 119

QY 119 VYGDVWMBRSTKR-HAGPFDLD-----RLLEFETN--LCHQSI-FYRRELFDGIGPNLYR 170
Db 120 AFGDVSRQKLEREHYWNVDLNNSESIESLENCYIAHPVMMKMSVLKALGGYNLYK 179
QY 171 VWADWDFNRCFNSPALITRYMDVVIS-----YNDMTGFS 206
Db 180 RTEDYNLWLR-----ATAKGYKIAMLEBKMKIRLHNSDKIHRDAEGFS 223
RESULT 7
US-10-282-122A-66207
; Sequence 66207, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66207
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-66207

Query Match 13.4%; Score 188; DB 15; Length 299;
Best Local Similarity 29.8%; Pred. No. 4.7e-12;
Matches 61; Conservative 33; Mismatches 91; Indels 20; Gaps 7;
QY 3 APVFSIIPTFNAAVTLQACLSIGSVGTQYREVEVLVDGGSTDRTLTDLTANSFRPELGSRL 62
Db 16 APLVSVVAPCFNAEKYLEALRSIYEQDYPNFVEIIVDGGSTDSYAMLEQLQKVHGSQL 75
QY 63 VVHSGPDGPDYDAMNRGVGATGWLFLGADDTLYEPTTLAQVAAPFLGDH-----AASH 117
Db 76 --YQQQNGVSAALNFGRLHARGDYVATPDLDDIIML-PHLSVRAAYLDQHPVEVCVAL 132
QY 118 LVYGDVWMBRSTKSRHAG---PFIDLRLFLFETNLCHQSI-FYRRELFDGIGPNLYRW 173

RESULT 9

US-10-282-122A-59156 ; Sequence 59156, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

Query Match	13.0%;	Score 182;	DB 15;	Length 291;
Best Local Similarity	23.5%;	Pred. No. 2.1e-11;		
Matches	61;	Conservative 49;	Mismatches 134;	Indels 16; Gaps 8;
Qy	4	PVFSIIPTFNAAVTTLQACIGSIVGQTYRVEVVLVDGSTDTRDTLDIANSFRPELGSRLV	63	
Db	5	PLVSVPVITHNRAELLSNAKSVINQYKIEIIICNDGSGDNTDDVVNKLQKEYKDIPF	64	
Qy	64	VHSGPDGDPYDA---MNRGVGVATGEVWLFLGADDTLYEPTTLAQVAFLG----	DHAAS 116	
Db	65	VYL-KNDSPWGAFCFSNRNCIAASGYTLTGLD-HDDYFLFNRIESLIKVCNERKIDLVCS	122	
Qy	117	HLVYGVDVWRSTKSRHAGFPDLRLLEFETNLCHQSIFYRRELFDGIGYNLYRVVWADWD	176	
Db	123	NLIFKDGKLRCKNNKNSGVITSDMGVE-NLVGNQLLTRLSYFEDVGGFDTNFPQAWQYD	181	
Qy	177	--FNIRCFSPALITRYMDVVISEYNDMTGFSMRQGTDFEKRKLPMYFWVAGHWETCRM	234	
Db	182	LWTRIIMKFGPCIKTNEATYVMDVDENDRKRISTSKAHIGYRSFIEKHSETLSEB--MKS	239	
Qy	235	LAFKLD--KENRRALRLRL	252	
Db	240	SLFIRDLINRNKKIPLKTL	259	

```

RESULT 10
US-10-282-122A-69114
; Sequence 69114, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert

```

```

; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69114
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-10-282-122A-69114

Query Match      13.0%; Score 182; DB 15; Length 327;
Best Local Similarity 20.8%; Pred. No. 2.5e-11;
Matches 70; Conservative 46; Mismatches 93; Indels 128; Gaps 13;

QY 1 MTAPVFSIIPTFNAVTLQACLSIGVQTVREVEVLVDGGSTDRDLDIANSFRPELGS 60
DB 1 MKTPIISVIVPFNEASRITLLESLLCQTFHDFEVIINDGSTDINSVDIAMLYCQQ-DN 59
QY 61 RLIVHSIGDDGPDYDAMRGVGVATGEWVLFLGADDTLYEPTTLA-----QVAAFLG 111
DB 60 RPHLYHQTNQGLSSARNTGLKYAQGDWIVFDSDDFI-KPQLLAHWHQLACEQHIDVLIG 118
QY 112 D-----HAASHLVYGDVVMRSTKSRHA 133
DB 119 NGERYDVHNPQKHQTIHQRPYQVVISQEWVHIAVTHQHPHFVWLQFIRHEILKXH- 177
QY 134 GPFDLRLLFETNLCHQSIFYRREL-----FDGIGPYNLRVYRWADWDFNRCFNSP 185
DB 178 -----HLRFIDGLYHEDILWTTQLALVTQIRIGFD-----DQPLYYYC-ANP 217
QY 186 ALITR-----YMDVVI-----SEYNDMTGFSNRQGTDXE-----FRKR 218
DB 218 NSITRKNPNQPKAKAQSILQIVLQLLQADKQHQPLLALRQALREFGSGFFILFRKR 277
QY 219 -----LPMYF-----WVAGWETCR 232
DB 278 CDTKNQTTIAQQFRYHYLFPALNKGANQRWRFICR 314

```

```

RESULT 11
US-10-282-122A-75381
; Sequence 75381, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangeu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert

```

```

; APPLICANT: Ohleen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75381
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-10-282-122A-75381

Query Match      13.0%; Score 182; DB 15; Length 344;
Best Local Similarity 22.4%; Pred. No. 2.7e-11;
Matches 70; Conservative 45; Mismatches 102; Indels 96; Gaps 10;

QY 7 SIIITPFNAVTLQACLSIGVQTVREVEVLVDGGSTDRDLDIANSFRPELGSRLVHS 66
DB 9 SIIIPLYNAGADFACNASLIAQTWSALEIIIVNDGSTDHSIETIAKHYAEHPHVRLHQ 68
QY 67 GPDGPDYDAMRGVGVATGEWVLFLGADDTLYEPTTLAQVAAFLGDHAAASHLVYGDVVMR 126
DB 69 A-NAGASVARNLGLQAATG DYVAFVDADLVYPKMYETLMTWALNDDLDVACQCNADWCVR 127
QY 127 STKSRHA-----GPFDL-----DRLLPET 145
DB 128 --KTCHAWQSITPDLRSTGVLSGPDWLRMALASRTHVVMGVYRREALITDNNITFVP 185
QY 146 NLCHQSIFYRRELFDGIGPYNLRVYRWADWDFNRCFNSPALITRYMDVITSEY---NDM 202
DB 186 GLHQDILWSTEVN-----FN-----ATRVRYTEQSLYKLYFLHDNS 221
QY 203 TGFSNRQGT-----DKEFRKELPMY---FWVAGWETCRRLAFLKCD 240
DB 222 VSRLOQGNKNLNYQRHYIKITRLLKLNRDYARRIPIYPPEFRQOITWEALRVCHAVRKE 281
QY 241 KENRRRLALRRLI 253
DB 282 PD---ILTRQRM 291

RESULT 12
US-10-282-122A-72785
; Sequence 72785, Application US/10282122A

```

Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Cart, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 72785
LENGTH: 343
TYPE: PRT
ORGANISM: Salmonella paratyphi A
US-10-282-122A-72785
Query Match 12.8%; Score 180; DB 15; Length 343;
Best Local Similarity 22.0%; Pred. No. 4.5e-11;
Matches 69; Conservative 46; Mismatches 102; Indels 96; Gaps 10;
QY 7 SIIIPFNAAVTLOACLSIGVGTREVEVVLVDGSGTDTLTDIANSFRPELGSRLVHS 66
DB 9 SIIIPLYNAGADFNACMSALIAQTWSALSIIVNGSDHSVEIAKHVAHYPHVRLHQ 68
QY 67 GPDDGPDYDAMRGVGVATGEWVFLGADDTLYEPTTLAQVAFLGDAASHLVYGDVVMR 126
DB 69 A-NAGASVARNLGLQAATGCVYVAFVADADDLVYPMYETLMTWALNDLDDVAQCNDWCVR 127
QY 127 STKSRA-----GPDFL-----DRLLPET 145
DB 128 --KTGHAWOSIPTDLRLSTGVLSGPDWLRMALASRRWTHVVMGVYRRLAIIDNNITFVP 185
QY 146 NLCHOSIFVRRLEFDGIGYNLRYRVWADWDNIRCFNSPALITRYMDVVISEY---NDM 202
DB 186 GLHQDILWSTEV-----FN-----ATRVRYTEOSLYKYLHNS 221
QY 203 TGFMSRQGT-----DKFRKRLPMY---FWVAGMETCRMLAFLKD 240
DB 222 VSLRQGNKNLNYQRHYIKITRLLEKLNDRYARRIPIYPEFRQITWEALRVCHAVRKE 281
QY 241 KENRLALRLI 253

Db 282 PD---ILTRQRM 291
RESULT 13
US-09-767-041-36
Sequence 36, Application US/09767041
Patent No. US20020055168A1
GENERAL INFORMATION:
APPLICANT: Smith, Hilda
TITLE OF INVENTION: STREPTOCOCCUS SUIS VACCINES AND DIAGNOSTIC TESTS
FILE REFERENCE: 2183-4726
CURRENT APPLICATION NUMBER: US/09/767,041
CURRENT FILING DATE: 2001-01-22
PRIOR APPLICATION NUMBER: PCT/NL99/00460
PRIOR FILING DATE: 1999-07-19
PRIOR APPLICATION NUMBER: EP98202465.5
PRIOR FILING DATE: 1998-07-22
PRIOR APPLICATION NUMBER: EP98202467.1
PRIOR FILING DATE: 1998-07-22
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.0
SEQ ID NO 36
LENGTH: 278
TYPE: PRT
ORGANISM: Streptococcus suis
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: CPS1K
US-09-767-041-36
Query Match 12.8%; Score 179; DB 9; Length 278;
Best Local Similarity 23.6%; Pred. No. 4.3e-11;
Matches 61; Conservative 47; Mismatches 93; Indels 59; Gaps 10;
QY 2 TAPVFSIIPTFNAAVTLOACLSIGVGTREVEVVLVDGSGTDTLTDIANSFRPELGSR 61
DB 3 TISKISIIIVPIYNVEKYSKICDSIVNQTXYKHIELLVNDGSDTNDSEECILAYAKK-DSR 61
QY 62 LVVHSGPDDGPDYDAMRGVGVATGEWVFLGADDTLY-----EPTTLAQVAAF 109
DB 62 IRYFKKENGSLSDARNYGISRAKGDYLAFLDSDDDFIHSEFIQRLHEALRENALVAVAGY 121
QY 110 LGDHAASHLVYGDVVMRSTKSRHAGFPDLDRLLFTNLIC-----HOSI-----FYRR 156
DB 122 DRVDASGHFLTAEP LPTN-----QAVLSGRNVCKKLEADGHRFVVACNKLYKK 170
QY 157 ELFDGIGYNLRY---RVWADWDNIRCFNSPALITRYMDVVISEYNDMTGSMRQG 210
DB 171 ELFE-----DFRFEKGKTHEDEYFTYRLLYELEKVAIVKCECLYYVVDRENSITSSM--- 222
QY 211 TDKEP-----RKRLPMY 222
DB 223 TDHREPHCLLEFQNERMDFY 241
RESULT 14
US-10-282-122A-67095
Sequence 67095, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Cart, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

```
; FILE REFERENCE: ELITRA_034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67095
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Pasteurella multocida
US-10-282-122A-67095

Query Match      12.8%; Score 179; DB 15; Length 477;
Best Local Similarity 21.6%; Pred. No. 9.2e-11;
Matches 59; Conservative 54; Mismatches 98; Indels 62; Gaps 7;

QY 1 MTAPVFSIIPTFNAAVTLOACLGSIYGQTYREVVEVLVDGSGDRTLDIANSFPRLGSGS 60
Db 211 LNAPLITILVTTFNSQKSIKNTLSNFQSYNIEIIVDHSQDNTWSILQAYTKYKN 270

QY 61 RLIVHSGPDGPDYDAMNRGVGVATGEWVLFGLGADDTLYEPTTTLAQVAAFLGDHAAASHLVY 120
Db 271 IKIISKENGVYVAKNIGLKYASGEFITCQSDDDWAHPQKALQVAPL-----LQH 322

QY 121 GDVVMRSTKSRHAGP-----FDLRLLEFETNLCHQSIPYRRELFDGIGPYNLRVY 171
Db 323 KELIVTFSKWVRLLDPIGNPYARTIYPLMLRPSS-----ALFRKKEVCE-----KTAL 370

QY 172 WADWDFNIRCFSPALITRYMDVVISYNDMTGFSMROGTDKEFRKRLPMYFWVAGWETC 231
Db 371 W-DW-----VRIGADSEFNARUKLIFGHKGYYTV 398

QY 232 RRLAFLKDKENRRRLALRTLIRVKAVERSA 264
Db 399 NKPLTFGAHRENSLMTAQSTGY-VNGVSLPREA 430

RESULT 15
US-09-767-041-22
; Sequence 22, Application US/09767041
; Patent No. US20020055168A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUITS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4726
; CURRENT APPLICATION NUMBER: US/09/767,041
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/NL99/00460
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: EP98202465.5
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: EP98202467.1
; PRIOR FILING DATE: 1998-07-22

; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Streptococcus suis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: CPS2K
US-09-767-041-22

Query Match      12.7%; Score 178; DB 9; Length 332;
Best Local Similarity 21.9%; Pred. No. 7.1e-11;
Matches 68; Conservative 53; Mismatches 100; Indels 90; Gaps 12;

QY 7 SIIITPFNAAVTLOACLGSIYGQTYREVVEVLVDGSGDRTLDIANSFPRLGSLVHVS 66
Db 5 SIIIVPIYVNEQYLSKCINSIVNQTYKHIEILLVNDGSDNSEEICLAVAKK-DSRIRYFK 63

QY 67 GPDDGPDYDAMNRGVGVATGEWVLFGLGADDTLY-----EPTTLAQVAAFLGDHA 114
Db 64 KENGGLSDARNYGISRAKGDYLAFLDSDDFIHSBFIQRLHEAIRRENALVAVAGYDRVDA 123

QY 115 ASHLVYGDVVMRSTKSRHAGPFDLRLLEFETNLC-----HOSI-----FYRRELFDG 161
Db 124 SCHFLTAEPPLPTN-----QAVLSGRNVCKLLEADGHRFVVAWNKLYKKELEFD- 171

QY 162 IGPYNLRY---RVWADWDFNIRC---FSNPALITRYMDVVISYNDMTGFSMROGTDKEP 215
Db 172 -----FRPEKGIHEDEVFTYRLLVELEKVAIVKECLVYVVDRENSIITSSM---TDHRF 223

QY 216 -----RKRLPMY-----FWVAGHETCERML---AF 237
Db 224 HCLLEFQNERMDFYESRGDKELLECYSRFLAFAVLFLGKYNHMLSKQCKKLQTLFRIVY 283

QY 238 LKDKENRRRLAL 248
Db 284 KQLKQNKRLAL 294

Search completed: February 27, 2005, 21:23:23
Job time : 136 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2005, 21:02:03 ; Search time 43 Seconds
(without alignments)
461.783 Million cell updates/sec

Title: us-10-805-311-24
Perfect score: 1401
Sequence: 1 MTAPEVFSIIITPFAAVTLQ.....ALRTRLIRKAVKERSGAEP 266

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	788	56.2	274	US-08-311-731A-16	Sequence 16, Appl
2	189.5	13.5	331	US-09-634-238-247	Sequence 247, App
3	188	13.4	303	US-09-252-991A-29155	Sequence 29155, A
4	178	12.7	340	US-09-543-681A-8334	Sequence 8334, Ap
5	174	12.4	330	US-09-134-000C-6392	Sequence 6392, Ap
6	172	12.3	333	US-09-107-532A-5123	Sequence 5123, Ap
7	164.5	11.7	842	US-09-134-000C-5749	Sequence 5749, Ap
8	163.5	11.7	534	US-09-134-000C-5087	Sequence 5087, Ap
9	163	11.6	721	US-09-107-532A-6889	Sequence 6889, Ap
10	161	11.5	324	US-08-597-236-10	Sequence 10, Appl
11	161	11.5	324	US-08-746-682A-10	Sequence 10, Appl
12	158	11.3	348	US-08-312-387B-3	Sequence 3, Appli
13	158	11.3	348	US-08-312-387B-11	Sequence 11, Appl
14	158	11.3	348	US-08-683-426-3	Sequence 3, Appli
15	158	11.3	348	US-08-683-426-11	Sequence 11, Appl
16	158	11.3	348	US-08-683-458-3	Sequence 3, Appli
17	158	11.3	348	US-08-683-458-11	Sequence 11, Appl
18	158	11.3	348	US-08-878-360-3	Sequence 3, Appli
19	158	11.3	348	US-08-878-360-11	Sequence 11, Appl
20	158	11.3	348	US-08-478-140B-3	Sequence 3, Appli
21	158	11.3	348	US-08-478-140B-8	Sequence 8, Appli
22	158	11.3	348	US-09-333-412-3	Sequence 3, Appli
23	158	11.3	348	US-09-333-412-11	Sequence 11, Appl
24	158	11.3	348	US-09-338-943-3	Sequence 3, Appli
25	158	11.3	348	US-09-338-943-8	Sequence 8, Appli
26	158	11.3	348	US-10-007-267A-3	Sequence 3, Appli
27	158	11.3	348	US-10-007-267A-11	Sequence 11, Appl

28	156.5	11.2	270	4	US-09-495-406-25	Sequence 25, Appl
29	156.5	11.2	270	4	US-09-816-028A-39	Sequence 39, Appl
30	156.5	11.2	270	4	US-10-303-162-39	Sequence 39, Appl
31	156.5	11.2	270	4	US-10-303-134-39	Sequence 39, Appl
32	155.5	11.1	376	4	US-09-902-540-13915	Sequence 13915, A
33	154.5	11.0	301	4	US-09-252-991A-33096	Sequence 33096, A
34	151.5	10.8	328	4	US-09-583-110-3554	Sequence 3554, Ap
35	151.5	10.8	331	4	US-09-107-433-4710	Sequence 4710, Ap
36	151	10.8	327	4	US-09-107-532A-6181	Sequence 6181, Ap
37	150	10.7	187	4	US-09-973-457-4	Sequence 4, Appli
38	149.5	10.7	703	4	US-09-902-540-11194	Sequence 11194, A
39	149.5	10.7	972	4	US-09-469-200E-9	Sequence 9, Appli
40	147.5	10.5	965	4	US-09-437-277-3	Sequence 3, Appli
41	147	10.5	750	4	US-09-489-039A-8940	Sequence 8940, Ap
42	146.5	10.5	301	4	US-09-495-406-15	Sequence 15, Appl
43	146.5	10.5	301	4	US-09-816-028A-27	Sequence 27, Appl
44	146.5	10.5	301	4	US-10-303-162-27	Sequence 27, Appl
45	146.5	10.5	301	4	US-10-303-134-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-08-311-731A-16
; Sequence 16, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/311,731A
; APPLICATION NUMBER: US/08/311,731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 274 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: MYCOBACTERIUM TUBERCULOSIS
; US-08-311-731A-16

Query Match 56.2%; Score 788; DB 4; Length 274;
Best Local Similarity 60.9%; Pred No. 1.5e-83;
Matches 156; Conservative 24; Mismatches 54; Indels 22; Gaps 2;
Qy 1 MTAPEVFSIIITPFAAVTLQACLSIGQTYREVEVVLVDGGSTDRTLDIANSFRPELGS 60

Db 19 MAAPNFSEIIPTLNVAAVLPACLDSIARQTCGDFELVLVDGGSTDTLIDIANIAPNLGE 78
Qy 61 RLNVHSGPDDGYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAAFLGDHAAASHLVY 120
Db 79 RLIIHRDTDQGYDAMNRGVDLATGTWLLFLGADDSLYEADTLARVAAFIGEHEPSDLVY 138
Qy 121 GDVWRSKSRHAGPFDRLIFETNLCHQSIQYREVEVLVDGGSTDRDLDIANSFRPELGSRLV 180
Db 139 GDVWRSKSRHAGPFDRLIFETNLCHQSIQYREVEVLVDGGSTDRDLDIANSFRPELGSRLV 198
Qy 181 CFSNPALITRYMDVVISYNDMTGSMRQGTDKFPRKRLPMYFWVAGWETCKRMLAFKLD 240
Db 199 CFSNPALITRYMDVVISYNDMTGSMRQGTDKFPRKRLPMYFWVAGWETCKRMLAFKLD 240
Qy 241 KENRLALRTRLIRVK 256
Db 239 --STRLGIRLVILVR 252

RESULT 2
US-09-634-238-247
; Sequence 247, Application US/09634238
; Patent No. 6544772

GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka J.
; APPLICANT: Bloksberg, Leonard, N.
; APPLICANT: Lubbers, Mark W.
; APPLICANT: Dekker, James
; APPLICANT: Christenson, Anna C.
; APPLICANT: Hölland, Ross
; APPLICANT: O'foole, Paul W.
; APPLICANT: Reid, Julian R.
; APPLICANT: Cbolbear, Timothy
; TITLE OF INVENTION: Polynucleotides, materials incorporating
; TITLE OF INVENTION: them and methods for using them.

FILE REFERENCE: 11000.1043UI
CURRENT APPLICATION NUMBER: US/09/634,238
CURRENT FILING DATE: 2000-08-08
NUMBER OF SEQ ID NOS: 422
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 247
LENGTH: 331
TYPE: PRT
ORGANISM: Lactobacillus rhamnosus
US-09-634-238-247

Query Match 13.5%; Score 189.5; DB 4; Length 331;
Best Local Similarity 28.2%; Pred. No. 2e-13; 93; Indels 27; Gaps 9;
Matches 62; Conservative 38; Mismatches 93; Indels 27; Gaps 9;
Qy 4 PVFSIIPTFNAAVTLQACLSIVGQTYREVEVLVDGGSTDRDLDIANSFRPELGSRLV 63
Db 7 PLVSIIVPIYVNEKYLQRCIDSLFAQTYVNIELVLVDGSDRSLSICK-NAQQDERV 65
Qy 64 VHSGBPDDGYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAAFL-GDHAASHLVYGD 122
Db 66 VFSKTNGVADTRNFGVSVARGWISFVDPDDYV-DPDYIYLYGILVKSNAWTMSICOHR 124
Qy 123 VVMRSK---SRHAGPFDRLIFETNLCHQSIQYREVEVLVDGGSTDRDLDIANSFRPELGSRLV 175
Db 125 NVYKNGKQTWLYEGPAVLDD-----SHTAV--KRLLYDD-----QIDTSWAKLYPAW 170
Qy 176 DFNIRCFSPALITRYMDVVISYNDMTGSMRQGTDKF 215
Db 171 VFKKIHPFKGRL---FEDIAATYKTFILASDSIAVGSSEAKY 207

RESULT 3
US-09-252-991A-29155
; Sequence 29155, Application US/09252991A
; Patent No. 6551795

GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29155
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29155

Query Match 13.4%; Score 188; DB 4; Length 303;
Best Local Similarity 29.8%; Pred. No. 2.6e-13;
Matches 61; Conservative 33; Mismatches 91; Indels 20; Gaps 7;
Qy 3 APVFSIIPTFNAAVTLQACLSIVGQTYREVEVLVDGGSTDRDLDIANSFRPELGSRL 62
Db 20 APLVSVVAPCFNAEKYLEEARLSIYEQDYPNFEVLIIVDDGSTDNSYAMLEQLQKVHGFQL 79
Qy 63 VHSGBPDDGYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAAFLGDH-----AASH 117
Db 80 --YRQONQGVSAALNFGLOHARGDYVATPDLDIIML-PHSLSVRAAYLDQHPVGCVGAL 136
Qy 118 LVYGDVVMRSKSRHAG---PFDLRLIFETNLCHQSI-FYRRELFDGIGPYNLRYRVA 173
Db 137 VIYDSGQETKQNGNRIQLDFYLLGNAYVCAPVSLYRMEALRAAGFYDPEIKV-Q 195
Qy 174 DWDFNIRCFSN-----PALITRY 191
Db 196 DFQMTLRIASQGYQIHKLPVLVTRY 220

RESULT 4
US-09-543-681A-8334
; Sequence 8334, Application US/09543681A
; Patent No. 6605709
GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8334
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-8334

Query Match 12.7%; Score 178; DB 4; Length 340;
Best Local Similarity 20.7%; Pred. No. 4.6e-12;
Matches 69; Conservative 46; Mismatches 91; Indels 128; Gaps 13;
Qy 4 PVFSIIPTFNAAVTLQACLSIVGQTYREVEVLVDGGSTDRDLDIANSFRPELGSRLV 63
Db 17 PIISVIVMFENASRIITLLESLLCQTFHDIFEVIINDGSTDNSVDIAMLYCQO-DNRPH 75
Qy 64 VHSGBPDDGYDAMNRGVGATGEWVFLGADDTLYEPTTLA-----QVAAFGLD-- 112
Db 76 LYHOTNOGLSSARNAGLUKYAQGDWIVFDSDDFI-KPOLLAHWHQLASEQUHIDVLIGNGE 134
Qy 113 -----HAASHLVYGDVVMRSKSRHAGPF 136

Db 135 RYDVNPKQHTIHOQPOYQOVISGEWVIHANTQHPHFWLQFIRHBIKKH----- 190
Qy 137 DLDRLLFETNLCHQSIFVRREL-----FDGIGPNLYRVWADWDFNRCFSNPALI 188
Db 191 ---HLRFIDGLYHEDILWTTLQALVTQRIGD-----DQPLYVC-ANFNSI 233
Qy 189 TR-----YMDVVI-----SEYNDMTGFSMRQGTDKB-----FRKR--- 218
Db 234 TRKNPQKEAKAQSYLQIVILQLLQADKQHQPLIALRQOALRELGSFPLFRKRCDT 293
Qy 219 -----LPMVF-----WVAGMETCR 232
Db 294 KQALIAQOQSTYHLFPALNKGAQNRORWFICR 327

RESULT 5

US-09-134-000C-6392
; Sequence 6392, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6392
; LENGTH: 330
; TYPE: PRN
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6392

Query Match 12.4%; Score 174; DB 4; Length 330;
Best Local Similarity 23.6%; Pred. No. 1.3e-11;
Matches 68; Conservative 49; Mismatches 99; Indels 72; Gaps 11;
Qy 4 PVFSIIITFNAAVTLQACLSIVGQTYREVVVLVDGSGTDRDLDIANSFRPELGSRLV 63
Db 8 PKSIIVPVYVNVKYLKCVRSILAQTFDFELILVDGSPDSSGAMCDQF-AEQDQKVK 66
Qy 64 VHSQPDGPDYDMNRGVGATGEWVFLGAD-----DTLYEPTTLAQVAAPLGDHAAHLV 119
Db 67 VIHKEGSLDARNAGIATGEVILGFVDSDDYIADMYE-----LL 108
Qy 120 YGDVVMRS-----TKSRHAGPDLDR---LIFETNLCH---QSIFVR 155
Db 109 YTNIVKEDADLSIGIYDVYEGKEPIVKSLIQGTFPREALLILQGNIIISVHAVNKLK 168
Qy 156 RELFCIGPNLYRVWADWDFNRCFSNPALITRYMDVISE-----YNDMTGFSMRQ 209
Db 169 RKLFP-----ADLRY---PGKYHEDSPFIVDLLSECQKVSIDSTQKYYHYHRMGSINTET 220
Qy 210 GTDKFRLKPLPMYVAGWETCRML-----AFKDKENRRLALRLLI 253
Db 221 FSDKQFE-----FIEAWERKELKKGAVIEAAHQVCFANFLV 261

RESULT 6

US-09-107-532A-5123
; Sequence 5123, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street

CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5123:
SEQUENCE CHARACTERISTICS:
LENGTH: 333 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...333
SEQUENCE DESCRIPTION: SEQ ID NO: 5123:
US-09-107-532A-5123

Query Match 12.3%; Score 172; DB 4; Length 333;
Best Local Similarity 27.5%; Pred. No. 2.3e-11; Indels 48; Gaps 6;
Matches 52; Conservative 33; Mismatches 56;
Qy 7 SIIPTFNAAVTLQACLSIVGQTYREVVVLVDGSGTDRDLDIANSFRPELGSRLV 66
Db 14 SIIIVPVYVNVKYLKCVDSILGQTFDFELLIDGSDTNSGICDEL-AKTDNRKIVH 72
Qy 67 GPDGPDYDMNRGVGATGEWVFLGADDTLYEPTTLAQVAAPLGDHAAHLVGDVVR 126
Db 73 KENGGLSDARNIGIEVAKGDFIGFIDSDVIDE-----DMYAFI-----YNNILKY 118
Qy 127 STKSRHAGPFD-----LDRLLFETNLCHQSIFVRRELFD 160
Db 119 DAELSMCGIYDVHVKKEIKLTPPSQLVTKEAIELVDGLVAVNAV--SKLYKELFE 176
Qy 161 GIGPNLYR 169
Db 177 -----NVRY 180

RESULT 7

US-09-134-000C-5749
; Sequence 5749, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15

QY 177 FNRFCFN-PALITRYMDVVVISEYNDMTGFSMRQGTDFK 215
DB 365 LVLRVTEQIPAEHIVHIDRVLYHWRITPGSTASNGEAKSY 404

RESULT 10

US-08-597-236-10
; Sequence 10, Application US/08597236
; Patent No. 5733765
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesca
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/597,236
; FILING DATE:
; CLASSIFICATION: 426
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci A., Allan
; REGISTRATION NUMBER: 30256
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 324 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-597-236-10

Query Match 11.5%; Score 161; DB 1; Length 324;
Best Local Similarity 35.9%; Pred. No. 4.2e-10;
Matches 42; Conservative 17; Mismatches 44; Indels 14; Gaps 3;
QY 5 VFSIIPTFNAAVTLQACLSIGVQTYREVEVLVDGGSTDRTLDIANSFRPELGSRLV 64
DB 6 LISIVIPVNVKYLEKCLQSQVQNTYNNFEVILVNDGSTDSSLSICEKFVNO-DKRFV 64
QY 65 HSGPDGPDYDANRQGVATGEWVFLGADDTLYEPTTLAQVAFLGHAASHLYVG 121
DB 65 FSKENGGMSSARNFGIKKAGSFITFVDSDDYI-----VKDYL-----SHLVAG 108

RESULT 11

US-08-746-682A-10
; Sequence 10, Application US/08746682A
; Patent No. 5786184
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesca
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York

; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,682A
; FILING DATE: 14-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/597,236
; FILING DATE: 20-JUN-1995
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci A., Allan
; REGISTRATION NUMBER: 30256
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 324 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-746-682A-10

Query Match 11.5%; Score 161; DB 1; Length 324;
Best Local Similarity 35.9%; Pred. No. 4.2e-10;
Matches 42; Conservative 17; Mismatches 44; Indels 14; Gaps 3;
QY 5 VFSIIPTFNAAVTLQACLSIGVQTYREVEVLVDGGSTDRTLDIANSFRPELGSRLV 64
DB 6 LISIVIPVNVKYLEKCLQSQVQNTYNNFEVILVNDGSTDSSLSICEKFVNO-DKRFV 64
QY 65 HSGPDGPDYDANRQGVATGEWVFLGADDTLYEPTTLAQVAFLGHAASHLYVG 121
DB 65 FSKENGGMSSARNFGIKKAGSFITFVDSDDYI-----VKDYL-----SHLVAG 108

RESULT 12

US-08-312-387B-3
; Sequence 3, Application US/08312387B
; Patent No. 5545553
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/312,387B
; FILING DATE: July 7, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 348 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-312-387B-3

Query Match 11.3%; Score 158; DB 1; Length 348;

Best Local Similarity 23.7%; Pred. No. 1e-09;

Matches 47; Conservative 37; Mismatches 80; Indels 34; Gaps 5;

```
QY 4 PVFSIIPTFNAAVTLQACLSIVGQTYREVEVVLVDGSGDRTLDIANSFRPELGSRLV 63
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 3 PLVSVLICAYNVEKYFAQSLAAVNVQWNRNDLILVDGSDGTGLAIADKFQKDSRIKI 62
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 64 VHSFGDDGPDYDAMNRGV-----GVATGEWVFLFGADD-----TLYEPTTLAQ 105
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 63 LAQANSGLIPSLNIGLDELAKSGGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 122
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 106 VAAFL-----GDHAASHLVYGVVWRSTKSRHAGFPDLDRLFLFETNLCHQSIFYRRE 157
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 123 MGAWLEVLSEKDGRLRHHKHGKIWKPKTRHEDIAAF-----PFPGNPINHNTMIMERS 178
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 158 LFDGIGPYNLYRVWADW 175
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 179 VIDG-----GLRYDTERDW 192
```

RESULT 13

US-08-312-387B-11

; Sequence 11, Application US/08312387B

; Patent No. 554553

; GENERAL INFORMATION:

; APPLICANT: Gotschlich, Emil C.

; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Klauber & Jackson

; STREET: 411 Hackensack Avenue

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/312.387B

; FILING DATE: July 7, 1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 600-1-095

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201 487-5800

; TELEFAX: 201 343-1684

; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 348 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-312-387B-11

Query Match

Best Local Similarity 11.3%; Score 158; DB 1; Length 348;

Matches 47; Conservative 37; Mismatches 80; Indels 34; Gaps 5;

```
QY 4 PVFSIIPTFNAAVTLQACLSIVGQTYREVEVVLVDGSGDRTLDIANSFRPELGSRLV 63
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 3 PLVSVLICAYNVEKYFAQSLAAVNVQWNRNDLILVDGSDGTGLAIADKFQKDSRIKI 62
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 64 VHSFGDDGPDYDAMNRGV-----GVATGEWVFLFGADD-----TLYEPTTLAQ 105
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 63 LAQANSGLIPSLNIGLDELAKSGGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 122
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 106 VAAFL-----GDHAASHLVYGVVWRSTKSRHAGFPDLDRLFLFETNLCHQSIFYRRE 157
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 123 MGAWLEVLSEKDGRLRHHKHGKIWKPKTRHEDIAAF-----PFPGNPINHNTMIMERS 178
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 158 LFDGIGPYNLYRVWADW 175
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 179 VIDG-----GLRYDTERDW 192
```

RESULT 14

US-08-683-426-3

; Sequence 3, Application US/08683426

; Patent No. 5705367

; GENERAL INFORMATION:

; APPLICANT: Gotschlich, Emil C.

; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Klauber & Jackson

; STREET: 411 Hackensack Avenue

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/683.426

; FILING DATE:

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/312,387

; FILING DATE: September 26, 1994

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 600-1-095B

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201 487-5800

; TELEFAX: 201 343-1684

; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 348 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-683-426-3

Query Match

Best Local Similarity 11.3%; Score 158; DB 1; Length 348;

Matches 47; Conservative 37; Mismatches 80; Indels 34; Gaps 5;

```
QY 4 PVFSIIPTFNAAVTLQACLSIVGQTYREVEVVLVDGSGDRTLDIANSFRPELGSRLV 63
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 3 PLVSVLICAYNVEKYFAQSLAAVNVQWNRNDLILVDGSDGTGLAIADKFQKDSRIKI 62
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 64 VHSFGDDGPDYDAMNRGV-----GVATGEWVFLFGADD-----TLYEPTTLAQ 105
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
```

Db 63 LAQAQNSGLIPSLNIGLDELAKSGGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 122
QY 106 VAAFL-----GHAASHLYVDVWMRSTKSRHAGPFDLDRLLFETNLCHQSIFYRRE 157
Db 123 MGAWLEVLSEEDGKRLARHHKHKWKPKTRHEDIAAF-----PFGNPIHNNTWIMRRS 178
QY 158 LFDGIGPNLYRYRVWADW 175
Db 179 VIDG-----GLRYDTERDW 192

RESULT 15
US-08-683-426-11
; Sequence 11, Application US/08683426
; Patent No. 5705367
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,426
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-683-426-11

Query Match 11.3%; Score 158; DB 1; Length 348;
Best Local Similarity 23.7%; Pred. No. 1e-09;
Matches 47; Conservative 37; Mismatches 80; Indels 34; Gaps 5;
QY 4 PVFSIIPTFNAAVTLQACLSIGVQTYREVEVLVDGGSTDRTLDIANSFPELGSRLV 63
Db 3 PLVSVLICAYNVEKYFAQSLAAVNTWRNLILIVDGGSTDTGTLAIKDFOKRDSRIKI 62
QY 64 VHSGLPDGPDYDAMRGV-----GVATGEWVFLGADD-----TLYEPTTLAQ 105
Db 63 LAQAQNSGLIPSLNIGLDELAKSGGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 122
QY 106 VAAFL-----GHAASHLYVDVWMRSTKSRHAGPFDLDRLLFETNLCHQSIFYRRE 157
Db 123 MGAWLEVLSEEDGKRLARHHKHKWKPKTRHEDIAAF-----PFGNPIHNNTWIMRRS 178
QY 158 LFDGIGPNLYRYRVWADW 175

Db 179 VIDG-----GLRYDTERDW 192
Search completed: February 27, 2005, 21:12:23
Job time : 45 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2005, 20:48:48 ; Search time 165 Seconds
(without alignments)
623.505 Million cell updates/sec

Title: US-10-805-311-24

Perfect score: 1401
Sequence: 1 MTAPEVSIITFTNAATLQ.....ALRTRLIRKAVKERSAEP 266

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: geneseqp1980s:*

2: geneseqp1990s:*

3: geneseqp2000s:*

4: geneseqp2001s:*

5: geneseqp2002s:*

6: geneseqp2003as:*

7: geneseqp2003bs:*

8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1401	100.0	266	2 AAW21775	Aaw21775 Protein e
2	1401	100.0	266	2 AAW21774	Aaw21774 Protein e
3	788	56.2	256	6 ABU34621	Abu34621 Protein e
4	788	56.2	274	7 ADB74267	Adb74267 Mycobacte
5	788	56.2	275	2 AAW21783	Aaw21783 Protein e
6	788	56.2	275	6 ABU36884	Abu36884 Protein e
7	298	21.3	297	2 AAW88309	Aaw88309 E. coli c
8	238.5	17.0	247	6 ABU50169	Abu50169 Protein e
9	205.5	14.7	298	4 AAB96313	Aab96313 Putative
10	202.5	14.5	262	2 AAW21779	Aaw21779 Protein e
11	201.5	14.1	321	5 ABP26805	Abp26805 Streptoco
12	192	13.7	336	6 ABU23911	Abu23911 Protein e
13	190	13.6	248	2 AAW88314	Aaw88314 Sugar tra
14	190	13.6	248	7 ADC01227	Adc01227 Enterohae
15	189.5	13.5	331	8 ADI67157	Adi67157 Lactobaci
16	188.5	13.5	333	6 ABU24245	Abu24245 Protein e
17	188	13.4	299	6 ABU38283	Abu38283 Protein e
18	188	13.4	303	7 ABO80409	Ab080409 Pseudomon
19	186.5	13.3	316	3 AAY43774	Aay43774 Amino aci
20	186.5	13.3	316	3 AAY54072	Aay54072 Enzyme EP
21	186	13.2	377	2 AAW21767	Aaw21767 Protein e
22	184.5	13.2	332	4 AAB47426	Aab47426 EpsN. 10/
23	182	13.0	291	6 ABU31232	Abu31232 Protein e
24	182	13.0	327	6 ABU41190	Abu41190 Protein e
25	182	13.0	344	6 ABU47457	Abu47457 Protein e

26	180.5	12.9	327	5 ABP26804	Abp26804 Streptoco
27	180	12.8	343	6 ABU44861	Abu44861 Protein e
28	179	12.8	278	3 AAY68976	Aay68976 CpsK pro
29	179	12.8	477	6 ABU39171	Abu39171 Protein e
30	178	12.7	340	7 ADF08049	Adf08049 Bacterial
31	177	12.6	421	6 ABU20309	Abu20309 Protein e
32	176	12.6	360	6 ABU20878	Abu20878 Protein e
33	175	12.5	334	3 AAY68963	Aay68963 Cps2K pro
34	174.5	12.5	333	5 ABB53519	Abb53519 Lactococc
35	174.5	12.5	623	5 ABB48553	Abb48553 Listeria
36	174.5	12.5	623	6 ABU32458	Abu32458 Protein e
37	174	12.4	330	7 ADH88507	Adh88507 Enterococ
38	173.5	12.4	331	6 ABU28458	Abu28458 Protein e
39	172.5	12.3	344	6 ABU28820	Abu28820 Protein e
40	172	12.3	333	7 ADC95496	Adc95496 E. faeciu
41	172	12.3	358	4 AAU34218	Aau34218 Staphyloc
42	172	12.3	573	4 AAU36881	Aau36881 Staphyloc
43	172	12.3	573	6 ABU15892	Abu15892 Protein e
44	172	12.3	574	6 ABM71992	Abm71992 Staphyloc
45	170.5	12.2	301	6 ABU02243	Abu02243 S. pneumo

ALIGNMENTS

RESULT 1

AAW21775

ID AAW21775 standard; protein; 266 AA.

AC AAW21775;

DT 17-OCT-2003 (revised)

DT 25-MAR-2003 (revised)

DT 11-MAR-1998 (first entry)

XX Protein encoded by ORF F of the GS region in M. paratuberculosis.

XX GS; pathogenecity island; pathogenic protein; mycobacterial disease;

KW cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;

KW vaccine; inflammatory disease; sarcoidosis; Johne's disease.

XX Mycobacterium avium subsp. paratuberculosis.

XX WO9723624-A2.

XX 03-JUL-1997.

XX 23-DEC-1996; 96WO-GB003221.

XX 21-DEC-1995; 95GB-00026178.

XX (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.

XX Hermondaylor J, Doran T, Millar D, Tizard M, Loughlin M, Sumar N;

XX Ford J;

XX WPI; 1997-351061/32.

XX N-PSDB; AAT74470.

XX New isolated pathogenicity island from mycobacteria - used to develop products for detection, diagnosis, prevention and treatment of mycobacteria infections.

XX Claim 1; Page 51; 62pp; English.

XX The present sequence represents a protein encoded by open reading frame (ORF) F of a novel polynucleotide sequence designated "GS". GS is a pathogenecity island of 8 kb of DNA comprising a core region of 5.75 kb with multiple ORFs and an adjacent transmissible element of 2.5 kb. The ORFs, and also the transmissible element, encode proteins which may be linked to pathogenecity, such as providing receptors for cellular recognition. GS was discovered and characterised using differential DNA analysis technology. It is found within Mycobacterium paratuberculosis

CC and it has also been identified in Mycobacterium avium subspecies
 CC silvaticum. These pathogenic mycobacteria cause chronic inflammation of
 CC the intestine and Crohn's disease in humans. The protein products of the
 CC ORFs of GS can be used for detecting mycobacteria or for diagnosing,
 CC treating or preventing mycobacterial disease. In particular they can be
 CC used as vaccines for inflammatory diseases such as Crohn's disease or
 CC sarcoidosis in humans or Johne's disease in animals. (Updated on 25-MAR-
 CC 2003 to correct PI field.) (Updated on 17-OCT-2003 to standardise OS
 CC field)
 XX Sequence 266 AA;
 SQ Query Match 100.0%; Score 1401; DB 2; Length 266;
 Best Local Similarity 100.0%; Pred. No. 1.3e-143; Indels 0; Gaps 0;
 Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTAPVFSIIPTFNAAVTLQACLSIVGQTYREVEVLVDGGSTDRTLDIANSFRPELGS 60
 DB 1 MTAPVFSIIPTFNAAVTLQACLSIVGQTYREVEVLVDGGSTDRTLDIANSFRPELGS 60
 QY 61 RLTVHSGPDDGPDYDAMNRGVGVATGEWVFLGADDTLYEPTTLAQVAAFLGDHAAASHLVY 120
 DB 61 RLTVHSGPDDGPDYDAMNRGVGVATGEWVFLGADDTLYEPTTLAQVAAFLGDHAAASHLVY 120
 QY 121 GDVVMRSTKSRHAGPFDLRLFFETNLCHOSIFYRRELFDGIGPNLYRVWADWDFNIR 180
 DB 121 GDVVMRSTKSRHAGPFDLRLFFETNLCHOSIFYRRELFDGIGPNLYRVWADWDFNIR 180
 QY 181 CFSNPALITRYMDVVISEYNDMTGFSMRQGTDKFPRKRLPMYFWVAGWETCRRMLAFLKD 240
 DB 181 CFSNPALITRYMDVVISEYNDMTGFSMRQGTDKFPRKRLPMYFWVAGWETCRRMLAFLKD 240
 QY 241 KENRRALRTRLIRVKAVERSAEP 266
 DB 241 KENRRALRTRLIRVKAVERSAEP 266
 RESULT 2
 AAW21774
 ID AAW21774 standard; protein; 266 AA.
 AC AAW21774;
 XX
 DT 17-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 11-MAR-1998 (first entry)
 XX
 DE Protein encoded by ORF F of GS region in M. avium subspecies silvaticum.
 XX GS; pathogenecity island; pathogenic protein; mycobacterial disease;
 KW cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
 KW vaccine; inflammatory disease; sarcoidosis; Johne's disease.
 XX Mycobacterium avium; subspecies silvaticum.
 OS
 XX WO9723624-A2.
 PN
 XX
 PD 03-JUL-1997.
 XX
 XX 23-DEC-1996; 96WO-GB003221.
 XX
 XX 21-DEC-1995; 95GB-00026178.
 XX
 XX (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.
 PA
 XX Hermonataylor J, Doran T, Millar D, Tizard M, Loughlin M, Sumar N;
 PI Ford J;
 XX
 XX WPI; 1997-351061/32.
 DR N-PSDB; AAT74469.
 XX
 XX New isolated pathogenecity island from mycobacteria - used to develop
 PT products for detection, diagnosis, prevention and treatment of

PT mycobacteria infections.
 XX Claim 1; Page 50; 62pp; English.
 XX
 CC The present sequence represents a protein encoded by open reading frame
 CC (ORF) F of a novel polynucleotide sequence designated "GS". GS is a
 CC pathogenecity island of 8 kb of DNA comprising a core region of 5.75 kb
 CC with multiple ORFs and an adjacent transmissible element of 2.5 kb. The
 CC ORFs, and also the transmissible element, encode proteins which may be
 CC linked to pathogenecity, such as providing receptors for cellular
 CC recognition. GS was discovered and characterised using differential DNA
 CC analysis technology. It is found within Mycobacterium paratuberculosis
 CC and it has also been identified in Mycobacterium avium subspecies
 CC silvaticum. These pathogenic mycobacteria cause chronic inflammation of
 CC the intestine and Crohn's disease in humans. The protein products of the
 CC ORFs of GS can be used for detecting mycobacteria or for diagnosing,
 CC treating or preventing mycobacterial disease. In particular they can be
 CC used as vaccines for inflammatory diseases such as Crohn's disease or
 CC sarcoidosis in humans or Johne's disease in animals. (Updated on 25-MAR-
 CC 2003 to correct PI field.) (Updated on 17-OCT-2003 to standardise OS
 CC field)
 XX Sequence 266 AA;
 SQ Query Match 100.0%; Score 1401; DB 2; Length 266;
 Best Local Similarity 100.0%; Pred. No. 1.3e-143; Indels 0; Gaps 0;
 Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTAPVFSIIPTFNAAVTLQACLSIVGQTYREVEVLVDGGSTDRTLDIANSFRPELGS 60
 DB 1 MTAPVFSIIPTFNAAVTLQACLSIVGQTYREVEVLVDGGSTDRTLDIANSFRPELGS 60
 QY 61 RLTVHSGPDDGPDYDAMNRGVGVATGEWVFLGADDTLYEPTTLAQVAAFLGDHAAASHLVY 120
 DB 61 RLTVHSGPDDGPDYDAMNRGVGVATGEWVFLGADDTLYEPTTLAQVAAFLGDHAAASHLVY 120
 QY 121 GDVVMRSTKSRHAGPFDLRLFFETNLCHOSIFYRRELFDGIGPNLYRVWADWDFNIR 180
 DB 121 GDVVMRSTKSRHAGPFDLRLFFETNLCHOSIFYRRELFDGIGPNLYRVWADWDFNIR 180
 QY 181 CFSNPALITRYMDVVISEYNDMTGFSMRQGTDKFPRKRLPMYFWVAGWETCRRMLAFLKD 240
 DB 181 CFSNPALITRYMDVVISEYNDMTGFSMRQGTDKFPRKRLPMYFWVAGWETCRRMLAFLKD 240
 QY 241 KENRRALRTRLIRVKAVERSAEP 266
 DB 241 KENRRALRTRLIRVKAVERSAEP 266
 RESULT 3
 ABU34621
 ID ABU34621 standard; protein; 256 AA.
 XX ABU34621;
 AC
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #20148.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX Mycobacterium bovis.
 OS
 XX WO20027183-A2.
 PN
 XX
 PD 03-OCT-2002.
 XX
 XX 21-MAR-2002; 2002WO-US009107.
 PF
 XX 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.


```
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX N-PSDB; ACA38491.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 62545; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 256 AA;
XX
XX Query Match 56.2%; Score 788; DB 6; Length 256;
XX Best Local Similarity 60.9%; Pred. No. 7e-77;
XX Matches 156; Conservative 24; Mismatches 54; Indels 22; Gaps 2;
XX
XX QY 1 MTAPEVSIIPFTNAAVTQACLSIGVQTVREVEVVLVDGSGDRTLDIANSFRPELGS 60
XX DB 1 MAAPMFSSIIPTLNVAAPLACLDLSIARQTCGDFELVLDGSGDRTLDIANSFRPELGS 60
XX
XX QY 61 RLVVHSGDPDGPYDAMNRGVATGEWVLFGLADDTLYEPTTLAQVAFLGDHAAHSLVY 120
XX DB 61 RLIIHRDQDQVYDAMNRGVDLATGTWLLFLGADDSLYEADTLARVAFAFGEHPSDLVY 120
XX
XX QY 121 GDVVMRSTKSRHAGPFDLRLFFETNLCHQSIYFRRELFDGIGPNLYRYRWADWDFNIR 180
XX DB 121 GDVIMRSTNFRWGGAFDRLDLFFKRNICHQAIYFRGLFGTIGPNLYRYRWADWDFNIR 180
XX
XX QY 181 CFSNPALITRYMVDVYSEYNDMTQFSMRQGTDKFRLKRLPMYFWVAGWETCRMLAFKLD 240
XX DB 181 CFSNPALVTRYMHWVWASYNFEGCLG-NTIVDKFRLKRLPM----- 220
XX
XX QY 241 KENRRLAIRTLIRVK 256
XX DB 221 --STRLGIRLVILVR 234
```

```
RESULT 4
ADB74267
ID ADB74267 standard; protein; 274 AA.
XX
XX AC ADB74267;
XX
XX DT 04-DEC-2003 (first entry)
XX
XX DE Mycobacterium tuberculosis non-naturally occurring peptide #15.
XX KW Non-naturally occurring peptide; anion pump protein; tuberculosis;
XX KW hypersensitivity reaction; tuberculostatic.
XX
XX OS Mycobacterium tuberculosis.
XX
XX PN US6593266-B1.
XX
XX PD 24-JUN-2003.
XX
XX PF 16-SEP-1994; 94US-00311731.
XX
XX PR 19-AUG-1993; 93US-00109181.
XX PR 22-OCT-1993; 93US-00142558.
XX
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX
XX PI Smith DR, Mao J;
XX
XX DR WPI; 2003-656441/62.
XX DR N-PSDB; ADB74252.
XX
XX PT New Mycobacterium tuberculosis anion pump peptide useful for as
XX PT tuberculosis vaccine and diagnosis of tuberculosis infection.
XX
XX PS Disclosure; SEQ ID NO 16; 26pp; English.
XX
XX CC The invention relates to a non-naturally occurring peptide of
XX CC Mycobacterium tuberculosis comprising an amino acid sequence
XX CC corresponding to an anion pump protein. The invention also relates to a
XX CC non-naturally occurring nucleic acid corresponding to a DNA sequence of
XX CC Mycobacterium tuberculosis or Mycobacterium leprae. The new peptide is
XX CC useful as a vaccine against Mycobacterium tuberculosis or Mycobacterium
XX CC leprae or for screening for new tuberculosis drugs. Purified proteins
XX CC derived from the sequences of the invention may elicit a specific immune
XX CC response. The peptide may also be used to detect hypersensitivity
XX CC reactions of individuals exposed to Mycobacterium tuberculosis or
XX CC Mycobacterium leprae. The proteins and peptides may be affixed to solid
XX CC supports to detect antibodies typical of hypersensitivity reactions, from
XX CC a patient's sera. This sequence represents Mycobacterium tuberculosis non
XX CC -naturally occurring peptide of the invention. Note: The sequence data
XX CC for this patent did not form part of the printed specification but was
XX CC obtained in electronic format directly from USPTO at
XX CC seqdata.uspto.gov/sequence.html.
XX
XX SQ Sequence 274 AA;
XX
XX Query Match 56.2%; Score 788; DB 7; Length 274;
XX Best Local Similarity 60.9%; Pred. No. 7e-77;
XX Matches 156; Conservative 24; Mismatches 54; Indels 22; Gaps 2;
XX
XX QY 1 MTAPEVSIIPFTNAAVTQACLSIGVQTVREVEVVLVDGSGDRTLDIANSFRPELGS 60
XX DB 19 MAAPMFSSIIPTLNVAAPLACLDLSIARQTCGDFELVLDGSGDRTLDIANSFRPELGS 78
XX
XX QY 61 RLVVHSGDPDGPYDAMNRGVATGEWVLFGLADDTLYEPTTLAQVAFLGDHAAHSLVY 120
XX DB 79 RLIIHRDQDQVYDAMNRGVDLATGTWLLFLGADDSLYEADTLARVAFAFGEHPSDLVY 138
XX
XX QY 121 GDVVMRSTKSRHAGPFDLRLFFETNLCHQSIYFRRELFDGIGPNLYRYRWADWDFNIR 180
XX DB 139 GDVIMRSTNFRWGGAFDRLDLFFKRNICHQAIYFRGLFGTIGPNLYRYRWADWDFNIR 198
```

Qy	181	CFSNPALITRYMDVWVISEYNDMTGFSMRQGTDKFPRKPLPMYFWVAGWETCRRMLAFKLD	240
Db	199	CFSNPALVTRYMHVVVSYNEFGGLS-NTIVDKFLKRLPM-----	238
Qy	241	KENRRLALRRLIRVK	256
Db	239	--STRLGIRLVIVLR	252
RESULT 5			
AAW21783	AAW21783 standard; protein; 275 AA.		
XX	AC	AAW21783;	
XX	DT	25-MAR-2003 (revised)	
DT	11-MAR-1998	(first entry)	
XX	DE	Protein encoded by ORF F from the GS region of M. tuberculosis.	
XX	GS	pathogenicity island; pathogenic protein; mycobacterial disease;	
KW	cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;		
KW	vaccine; inflammatory disease; sarcoidosis; Johne's disease; ss.		
XX	OS	Mycobacterium tuberculosis.	
XX	PN	WO9723624-A2.	
XX	PD	03-JUL-1997.	
XX	23-DEC-1996;	96WO-GB0003221.	
XX	21-DEC-1995;	95GB-00026178.	
XX	(SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.		
PA	Hermontaylor J, Doran T, Millar D, Tizard M, Loughlin M, Sumar N;		
PI	Ford J;		
XX	WPI; 1997-351061/32.		
DR	N-PSDB; AAT74477.		
XX	New isolated pathogenicity island from mycobacteria - used to develop		
PT	products for detection, diagnosis, prevention and treatment of		
PT	mycobacteria infections.		
XX	Claim 1; Page 56; 62pp; English.		
XX	The present sequence represents the protein encoded by open reading frame		
CC	(ORF) F, from M. tuberculosis. This ORF F has been found to have homology		
CC	with ORF F of a novel polynucleotide sequence designated "GS". GS is a		
CC	pathogenicity island of 8 kb of DNA comprising a core region of 5.75 kb		
CC	with multiple ORFs and an adjacent transmissible element of 2.5 kb. The		
CC	ORFs, and also the transmissible element, encode proteins which may be		
CC	linked to pathogenicity, such as providing receptors for cellular		
CC	recognition. GS was discovered and characterised using differential DNA		
CC	analysis technology. It is found within Mycobacterium paratuberculosis		
CC	and it has also been identified in Mycobacterium avium subspecies		
CC	silvaticum. These pathogenic mycobacteria cause chronic inflammation of		
CC	the intestine and Crohn's disease in humans. The protein products of the		
CC	ORFs of GS can be used for detecting mycobacteria or for diagnosing,		
CC	treating or preventing mycobacterial disease. In particular they can be		
CC	used as vaccines for inflammatory diseases such as Crohn's disease or		
CC	sarcoidosis in humans or Johne's disease in animals. (Updated on 25-MAR-		
CC	2003 to correct PI field.)		
XX	Sequence 275 AA;		
Qy	Match	56.2%; Score 788; DB 2; Length 275;	
Best	Local Similarity	60.9%; Pred. No. 7.8e-77;	
Matches	156; Conservative	24; Mismatches 54; Indels 22; Gaps 2;	

10

Query Match 14.4%; Score 201.5; DB 5; Length 321;
Best Local Similarity 30.1%; Pred. No. 7.4e-13;
Matches 52; Conservative 33; Mismatches 63; Indels 25; Gaps 6;

QY 7 SIIITTFNAVTLQACLSIGQTVREVEVLVDGGSTDRITLDIANSFPELGSRLVWHS 66

Db 5 SIIIPVNVQFLNECIESVLQAQTSYLNLEIILVNDGSTDNGDIDC-YYSIDGRIFVFH 63
Qy 67 GPDDGPDYAMRGVGVATGEWVLFAGADDTLYEPTTLAQVAFLGDHAAHLVYG----- 121
Db 64 KNGGLSDARNYGISRATGDIYLLDSDDYLYKEDAIERMVEP-SEKYNISVILGCVYEK 122
Qy 122 -----DVVMSRSTKSRHAGP-----FOLDRLLPETNLCQHSIFVRRRLF 159
Db 123 REQHIIINIVLEDEMIETISPVQAQNIYNDAYRAIF--TVAHNKL-YKRELF 172

RESULT 12

ABU23911
ID ABU23911 standard; protein; 336 AA.

XX AC ABU23911;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #9438.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Clostridium acetobutylicum.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI; 2003-029926/02.

XX DR N-PSDB; ACA27781.

XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 51835; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the
XX CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ

Sequence 336 AA;

Query Match 13.7%; Score 192; DB 6; Length 336;

Best Local Similarity 23.6%; Pred. No. 8.6e-12;

Matches 65; Conservative 57; Mismatches 113; Indels 40; Gaps 10;

Qy 4 PVFSIIITFNAAVTTLQACLGSIYQTYREVVEVLVDGSGTDRDLTIANSPFELGRLV 63
Db 5 PKVSIIVPVNYSERYLAETIESILDQTYNDPEFIIVDDGSTDESYNISSYANK-DNRII 63

Qy 64 VHSQPDGPDYAMRGVGVATGEWVLFAGADDT-----TLYEPTTLAQVAFLGDHAA 115
Db 64 VISREHRLGLVDSLNEGNIARIGKYIARMADADDISINNRIEKQFEFLNLDVDILGTRIE 123

Qy 116 SHLVYGVVMEST---KSRHAGPFDLRL--LPETN--LCHOSIFYRRELFDGIGPYNLR 168
Db 124 A---FGDIDKQKTIYNSAFSIFKDSQNIQOVFLTSCAIPHPSPVFKDSIVKLGRTKE 180

Qy 169 YRVWADWDFNIRCFSPALITRYMDVVISEYN-----DMTGFMSRQG---- 210
Db 181 YDTAEDYDLMLRAIRNGYKIVR-MDECLIKYRVHNSKTAVEMPNKMWVETMKAKIDYI 239

Qy 211 TDKEFRKRLPMYFWA--GMETCRMLAFLKDKEN 243
Db 240 NDTNKKDKVDYLMGASTGGKLVKKVVESTTDKEN 274

RESULT 13

AAW88314
ID AAW88314 standard; protein; 248 AA.

XX AC AAW88314;

XX DT 26-APR-1999 (first entry)

XX DE Sugar transferase involved in O157 antigen biosynthesis.

XX KW O antigen; O157 antigen; sugar transferase; wbdO gene; diarrhoea;
XX KW haemorrhagic colitis; diagnosis.

XX OS Escherichia coli.

XX PN WO9850531-A1.

XX PD 12-NOV-1998.

XX PF 01-MAY-1998; 98WO-AU0000315.

XX PR 01-MAY-1997; 97AU-00006545.

XX PR 22-JUL-1997; 97AU-00008162.

XX PA (UNSY) UNIV SYDNEY.

XX PI Reeves PR, Wang L;

XX DR WPI; 1999-059669/05.

XX DR N-PSDB; AAX06749.

XX PT Nucleic acid molecules specific for bacterial polysaccharide antigens -
XX PT useful for detecting specific strains in, e.g. food, faeces or patient
XX PT samples.

XX PS Disclosure; Fig 8; 165pp; English.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2005, 19:38:19 ; Search time 3386 Seconds
(without alignments)

9004.572 Million cell updates/sec

Title: US-10-805-311-23

Perfect score: 801

Sequence: 1 atgactgcgcagtggtctc.....aacgaagcgacgaaccgtag 801

Scoring table: IDENTITY NUC

Gapop 10_0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hc:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gsa1:*

9: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	49.4	6.2	1266	BZ558529	BZ558529 pa98401.2
2	48	6.0	751	CF867014	CF867014 trico09xd
3	48	6.0	805	CB897020	CB897020 trico09xd
C 4	46	5.7	1212	BZ560978	BZ560978 pacs2-164
5	45.6	5.7	660	BZ891972	BZ891972 Hml1_0147
6	44.8	5.6	558	CK308368	CK308368 hzma0_005
C 7	44.4	5.5	548	CK146970	CK146970 ZMBB0500
8	44.2	5.5	833	CK195343	CK195343 FGAS00378
9	43.4	5.4	491	CA026025	CA026025 HZ53N24r
10	43.4	5.4	658	BI960016	BI960016 HVSMen002
11	43.4	5.4	869	BI959554	BI959554 HVSMen002
12	43.2	5.4	608	AJ274182	AJ274182 AJ274182
13	43	5.4	802	CK201078	CK201078 FGAS00959
14	42.4	5.3	464	CA025392	CA025392 HZ52A13r
15	42.4	5.3	553	CA023059	CA023059 HZ45B23r
C 16	42	5.2	516	AA835320	AA835320 ak65a10.8
17	41.8	5.2	631	BI958865	BI958865 HVSMen001
18	41.6	5.2	963	CF885491	CF885491 trico082xi
19	41.2	5.1	1026	BE039820	BE039820 OC08F05 O
20	40.8	5.1	399	AV423282	AV423282 AV423282
C 21	40.8	5.1	445	BZ337760	BZ337760 ia90g07_b
22	40.8	5.1	661	CA092574	CA092574 SCVPM206
C 23	40.8	5.1	743	COS30618	COS30618 3530_1_20
24	40.8	5.1	868	BZ404229	BZ404229 OGAB507C

C 25	40.8	5.1	920	9	CG445414	CG445414 OG5D234TC
26	40.8	5.1	936	9	CG337316	CG337316 OGXC552TV
C 27	40.8	5.1	951	9	CG337305	CG337305 OGXC552TH
28	40.6	5.1	688	2	BF430504	BF430504 OG01F05T3
29	40.6	5.1	759	6	CB643463	CB643463 OSJNB04C
30	40.6	5.1	779	6	CB642703	CB642703 OSJNB03A
C 31	40.6	5.1	897	2	BE040583	BE040583 OF03D08 O
C 32	40.4	5.0	822	8	CG329310	CG329310 OGWL041TV
C 33	40.4	5.0	832	8	CG330216	CG330216 OGRAL86TV
C 34	40.4	5.0	832	9	CG443515	CG443515 OGVGM64TH
C 35	40.4	5.0	925	9	CL976984	CL976984 OGrifCC043
36	40.4	5.0	1029	9	CD913580	CD913580 GS50.118G
37	40.2	5.0	696	6	CA182349	CA182349 SCBGST311
C 38	40.2	5.0	710	6	BM377275	BM377275 EBem05_SQ
C 39	40	5.0	509	4	CF875058	CF875058 trico37xn
C 40	40	5.0	735	7	CB904315	CB904315 trico37xn
C 41	40	5.0	789	6	CK195936	CK195936 FGAS00438
42	40	5.0	835	7	BE608463	BE608463 sq33e11.y
C 43	39.8	5.0	463	2	BI947516	BI947516 HVSMEL000
44	39.8	5.0	505	4	BE969846	BE969846 601679418
45	39.8	5.0	642	2		

ALIGNMENTS

RESULT 1
BZ558529/c 1266 bp DNA linear GSS 17-DEC-2002
LOCUS pa98401_220.xl pacs2-164 Pseudomonas aeruginosa genomic clone
DEFINITION pa98401_220, Genomic survey sequence.
ACCESSION BZ558529
VERSION BZ558529.1 GI:27173107
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1266)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
source
1..1266
Location/Qualifiers
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone_lib="pacs2-164"
/clone="pa98401_220"
/note="clinical isolate 2-164 Whole genomic shotgun library."

Query Match 6.2%; Score 49.4; DB 8; Length 1266;
Best Local Similarity 57.1%; Pred. No. 0.0076;
Matches 109; Conservative 0; Mismatches 81; Indels 1; Gaps 1;
QY 4 ACTGGCCAGTGTTCGATATTCCTACCTTCATGCGCGGTGACGCTGCAAGCC 63
Db .259 AGCGAGCCATGGTTTCGTCATCATCGCTTCCTATACCGAGAAAGTACATAGAAGG 200
QY 64 TGCCTCGGAAGCATCGTCGGGACACCTACCGGAAGTGAAGTGGTCTCTTGTGACGCGC 123

Db 199 ACATCGCCAGCGTGTCTCAGACATACCCAGTGGAACTCTGCTGTGTGACGAT 141
 QY 124 GGTTCGACCGATCGGACCTCGACATCGCGAAGAGTTTCCGCCCGGAACTCGGCTCGGA 183
 Db 140 GGTTCGACCGATCGGACGCTCGAGCTTTTCCGCCCGGCTCCAGCGGAAGTACGCTTCGAC 81
 QY 184 CTGGTCGTTCA 194
 Db 80 CTCAGGCTCCA 70

RESULT 2
 CF867014 751 bp mRNA linear EST 31-OCT-2003
 LOCUS
 DEFINITION
 Hypocrea jecorina cDNA clone trico09xd01, Version 6 October 2003
 ACCESSION
 VERSION
 SOURCE
 ORGANISM
 Hypocrea jecorina (anamorph: Trichoderma reesei)
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

REFERENCE
 AUTHORS
 Diener, S.E., Dunn-Coleman, N., Foreman, P., Houfek, T.D.,
 Ward, M. and Dean, R.A.
 Teunissen, P.J.M., van Solingen, P., Dankmeyer, L., Mitchell, T.K.,
 in a comprehensive set of expressed sequence tags from Trichoderma
 reesei

JOURNAL
 COMMENT
 FEMS Microbiol. Lett. 230 (2), 275-282 (2004)
 Contact: Ralph A. Dean
 Fungal Genomics Laboratory
 North Carolina State University
 Campus Box 7251, Raleigh, NC 27695, USA
 Tel: 919-513-0020
 Fax: 919-513-0024
 Email: ralph_dean@ncsu.edu
 Seq primer: LT-F1 primer.

FEATURES
 source
 1..751
 Location/Qualifiers
 /organism="Hypocrea jecorina"
 /mol_type="mRNA"
 /strain="QM6a"
 /db_xref="taxon:51453"
 /clone="trico09xd01"
 /dev_stage="mycelia"
 /clone_lib="T.reesei mycelial culture, Version 6 October 2003"
 /note="Vector: pREP3Y; Site_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN
 Query Match 6.0%; Score 48; DB 7; Length 751;
 Best Local Similarity 54.0%; Pred. No. 0.018;
 Matches 121; Conservative 0; Mismatches 100; Indels 3; Gaps 1;
 QY 74 GCATCGTCGGGACGACCTACCGGGAAGTGGAAAGTGGTCTTCTGCGACGGCGGTTGACCG 133
 Db 206 GCATCTTCACAGAGCACCACCTCGATTGGAACTCATCATCTGCGACGAGGCTCGCCG 265
 QY 134 ATCGGACCTCGACATCGCGAAGAGTTTCCGCCCGGAACTCGGCTCGGACTGGTCGTTTC 193
 Db 266 ACGGACCCCAAGAGCTGCGCAACCACTCGTCAAGGCTTACGCCCCCAACGTCGTCCTCA 325
 QY 194 ACAGCGGGCCCGATGAT---GGCCCTACGACGATGAACCGCGGCTCGGCGTAGCCA 250
 Db 326 AGACTCGCTCCGCAAGCTGGGCGCTCGGCACCGCTTACGTCACGCGCTCGCATGTCGTA 385
 QY 251 CAGGCGAATGGGTACTTTTATAGGCGCGGACGACACCTCTTAC 294

Db 386 CGGGCAACTTCGTATCATCATGAGCGCGACTTTCAGCCACCAC 429
 RESULT 3
 CB897020 805 bp mRNA linear EST 02-JUL-2003
 LOCUS
 DEFINITION
 trico09xd01 T.reesei mycelial culture, Version 3 april Hypocrea
 jecorina cDNA clone trico09xd01, mRNA sequence.
 ACCESSION
 VERSION
 SOURCE
 ORGANISM
 Hypocrea jecorina (anamorph: Trichoderma reesei)
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

REFERENCE
 AUTHORS
 1 (bases 1 to 805)
 Foreman, P.K., Brown, D.B., Dankmeyer, L., Dean, R., Diener, S.,
 Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,
 Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,
 Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.
 Transcriptional regulation of biomass-degrading enzymes in the
 filamentous fungus Trichoderma reesei
 J. Biol. Chem. 278 (34), 31988-31997 (2003)

JOURNAL
 MEDLINE
 PUBMED
 COMMENT
 Contact: Pamela K. Foreman
 Genencor Intl.
 925 Page Mill Road, Palo Alto, CA 94304, USA
 Tel: (650) 846-7635
 Fax: (650) 621-7817
 Email: Pforeman@genencor.com
 Seq primer: LT-F1 primer.
 Location/Qualifiers
 1..805
 /organism="Hypocrea jecorina"
 /mol_type="mRNA"
 /strain="QM6a"
 /db_xref="taxon:51453"
 /clone="trico09xd01"
 /dev_stage="mycelia"
 /clone_lib="T.reesei mycelial culture, Version 3 april"
 /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

FEATURES
 source
 1..805
 Location/Qualifiers
 /organism="Hypocrea jecorina"
 /mol_type="mRNA"
 /strain="QM6a"
 /db_xref="taxon:51453"
 /clone="trico09xd01"
 /dev_stage="mycelia"
 /clone_lib="T.reesei mycelial culture, Version 3 april"
 /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN
 Query Match 6.0%; Score 48; DB 6; Length 805;
 Best Local Similarity 54.0%; Pred. No. 0.018;
 Matches 121; Conservative 0; Mismatches 100; Indels 3; Gaps 1;
 QY 74 GCATCGTCGGGACGACCTACCGGGAAGTGGAAAGTGGTCTTCTGCGACGGCGGTTGACCG 133
 Db 260 GCATCTTCACAGAGCACCACCTCGATTGGGAACCTCATCATCTGCGACGAGGCTCGCCG 319
 QY 134 ATCGGACCTCGACATCGGGAACAGTTTCCGCCCGGAACTCGGCTCGGAGCTGGTCGTTTC 193
 Db 320 ACGGACCCCAAGAGCTGCGCAACCACTCGTCAAGGCTTACGCCCCCAACGTCGTCCTCA 379
 QY 194 ACAGCGGGCCCGATGAT---GGCCCTACGACGATGAACCGCGGCTCGGCGTAGCCA 250
 Db 380 AGACTCGCTCCGCAAGCTGGGCGCTCGGCACCGCTTACGTCACGCGCTCGCATGTCGTA 439
 QY 251 CAGGCGAATGGGTACTTTTATAGGCGCGGACGACACCTCTTAC 294
 Db 440 CGGGCAACTTCGTATCATCATGAGCGCGCACTTCAGCCACCAC 483

RESULT 4
 BZ560978/c 1212 bp DNA linear GSS 17-DEC-2002
 LOCUS
 DEFINITION
 pacs2-164_2973.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
 pacs2-164_2973, genomic survey sequence.
 ACCESSION
 BZ560978

/organism="Hypocrea lixii"
/mol_type="mRNA"
/db_xref="taxon:5544"
/dev_stage="mycelium"
/lab_host="E.coliDH10"
/clone_lib="Trichoderma harzianum mycelium cDNA library"
/note="Vector: pBLUESCRIPT II SK+"

ORIGIN

Query Match 5.6%; Score 44.8; DB 7; Length 558;
Best Local Similarity 53.1%; Pred. No. 0.14;
Matches 119; Conservative 0; Mismatches 102; Indels 3; Gaps 1;

QY 74 GCATCGTGGGAGACCTACCGGAGTGGTCTTCGACGCGGTTGACCG 133
Db 128 GCACCTTTACAGAGAACCACTGGATTTGGAACTCATCATCGTACGACGCTCACCG 187

QY 134 ATCGACCCCTCGACATCGGACAGTTTCGCGCGGAACTCGGCTCGGACTGTCGTTTC 193
Db 188 ATGGACACAGACGTGGCCACACAGCTCGTCAAGCCCTACGCCCCACAGCTCGTCTCA 247

QY 194 ACAGCGGGCGCATGAT---GGCCCTACGACGCGCATGAACCGCGGCGTTCGCGTAGCCA 250
Db 248 AGACTCGATCCGCAAGCTGGGCTCGGAACGCGCTACGTCACGCGCTGCAGTTTGTCA 307

QY 251 CAGGCGAATGGTACTTTTTTAGGCGCGGACGACACCTCTAC 294
Db 308 CGGGCAACTTTGTCATCATCATGGAGCGCGACTTCAGCCACAC 351

RESULT 7
CC146970/c
LOCUS ZM5BB0009F05.r ZM5BB Zea mays genomic clone ZM5BB0009F05 3',
DEFINITION genomic survey sequence.
ACCESSION CC146970
VERSION CC146970.1 GI:30092163
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays

REFERENCE
AUTHORS Yu, Y., Kim, H. R., Hatfield, J., Soderlund, C., Bharti, A. K., Messing, J. and Wing, R.

TITLE Sequencing of the maize genome
JOURNAL Unpublished (2003)
COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: T7
BACKWARD: M13r
Plate: 0009 row: F column: 05
Seq primer: M13r
Class: BAC ends.

FEATURES
source
1..548
Location/Qualifiers
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ZM5BB0009F05"
/lab_host="DH10B"
/clone_lib="ZM5BBB"
/note="Vector: pBelOBAC11; Site 1: HindIII; Site 2: HindIII; Zea mays L. ssp. mays"

ORIGIN

Query Match 5.5%; Score 44.4; DB 8; Length 548;
Best Local Similarity 47.2%; Pred. No. 0.18;
Matches 135; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 46 GCGGTGACGCTGCAAGCCTCGCTCGAAGCATCGTCGGGCGACCTACCGGAAGTGGA 105
Db 432 GCGCTAACCGCGCGGCTGGGTTCGGGCGCTCAGGCGCCAGGGCTCCACGACGGCGG 373

QY 106 GTGCTCTTGTGACGCGGTTTCGACCGATCGGACCCCTCGACATCGCAACAGTTTCCGC 165
Db 372 GTCGCGCTTGGCTCGCGGAGCGGACTTGACAAAGTCGAGGAGCGGTGGCTCACCTC 313

QY 166 CGGAACTCGGCTCGGACTGCTGTTTCACAGCGGCGCGATGATGGCCCTTACACGCC 225
Db 312 CTTGAGTAGGTCTGACGCGCTCGTTCCCTCTCCACGAGGCGCGCGCTCCCGCG 253

QY 226 ATGAACCGCGCTCGGCGTAGCCACAGCGCAATGGGTACTTTTTTAGGCGCGCACGAC 285
Db 252 AGCGCCACAGCGTTCGAAGGCTCGGTCTCGATGCGCGCGCGCGGCTCGCGTCCGC 193

QY 286 ACCCTCTACGAACCAACCAACCGTTGCGCCAGGTAGCGCTTTTCTCG 331
Db 192 GACCGGACGGCGCCATAGCGTTGCAAGGACGCGTGTCTCCCGCG 147

RESULT 8
CK195343
LOCUS FGAS003782 Triticum aestivum FGAS: Library 3 Gate 6 Triticum
DEFINITION aestivum cDNA, mRNA sequence.
ACCESSION CK195343
VERSION CK195343.1 GI:39557733
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum

REFERENCE
AUTHORS Allard, P., Crosby, W. L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D., Genswein, B., Graf, R., Gulick, P., Hrycan, L. D., Laroché, A., Links, M. G., McCarthy, E. L., Monroy, A., Muzak, I., Nilsson, D., Peniket, C., Roach, J. L. and Sarhan, F.

TITLE Functional Genomics of Abiotic Stress in Wheat and Canola Crops
JOURNAL Unpublished (2003)
COMMENT Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033
Email: fgas.est@usask.ca

This sequence is the direct result of the Base calling software Phred (default parameters). It is the raw base calls. To aid in the identification of the high quality insert the software Lucy (default parameters) has been run on this sequence. Lucy identified the region [102,703].
Plate: L3C104 row: H column: 14.
Location/Qualifiers
1..833
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone_lib="Triticum aestivum FGAS: Library 3 Gate 6"
/note="Organ: Root; Vector: pCMV.SPORT6; Library tissue from Norstar. 7 mRNA populations were combined before constructing the library; 7 day non-acclimated roots, 1, 23, and 53 days cold-acclimated at 4C, and 30 minutes, 3 hours and 6 hours treated roots with 200mM NaCl.

FEATURES
source
1..833
Location/Qualifiers
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone_lib="Triticum aestivum FGAS: Library 3 Gate 6"
/note="Organ: Root; Vector: pCMV.SPORT6; Library tissue from Norstar. 7 mRNA populations were combined before constructing the library; 7 day non-acclimated roots, 1, 23, and 53 days cold-acclimated at 4C, and 30 minutes, 3 hours and 6 hours treated roots with 200mM NaCl.

Non-acclimated and cold-acclimated plants were grown in vermiculite while salt stressed plant were grown hydroponically. First strand synthesis in this library was done in the presence of methylated dCTP thereby protecting from internal cleavage with NotI."

ORIGIN

Query Match 5.5%; Score 44.2; DB 7; Length 833;
Best Local Similarity 50.2%; Pred. No. 0.22;
Matches 106; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 47 CGGTGAGCTGCGAGCTGCTCGAGAGCATGTCGGGCGAGACCTACCGGGAAGTGGAG 106
DB 547 CGGTGAGCTGCGAGCTGCTCGAGAGCATGTCGGGCGAGACCTACCGGGAAGTGGAG 606
QY 107 TGGTCTTCTGCGAGCGGCTTGCACCGATCGGACCTCGACATCGGACAGTTCCTCGGC 166
DB 607 CCTACGAGAGATGNCACCGCCACACATCGGAGCATCAACCTCATACCGTGGCC 666
QY 167 CGGAATCGGCTCGGACCTGCTGTTTCACAGCGGCGCCGATGATGCCCTTACGAGCCA 226
DB 667 AGAAGCTCTGGATCCCGCTGCCCTGCGAGCTGCGAGCCCGCGCGCGCGAGCTCT 726
QY 227 TGAACCGCGGCTCGGCTGAGCCACAGGCGA 257
DB 727 TGCACCTCGGCACATCGTCGAGCGCGGGA 757

RESULT 9

CA026025
LOCUS
DEFINITION
HZ53N24r HZ Hordeum vulgare subsp. vulgare cDNA clone HZ53N24
5-PRIME, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Hordeum vulgare subsp. vulgare
Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.

REFERENCE

AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 491)
Radchuk, V., Zhang, H., Weschke, W., Potokina, E. and Wobus, U.
Barley ESTs from developing seeds
Unpublished (2002)
Contact: Stein Nils

Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595

Email: stein@ipk-gatersleben.de
Insert Length: 491 Std Error: 0.00
Plate: 53 row: N column: 24
Seq primer: M3rev.

FEATURES

source

1. .491
Location/Qualifiers
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="barke"
/sub_species="vulgare"
/db_xref="GABI:279451"
/db_xref="taxon:112509"
/clone="HZ53N24"
/tissue_type="pericarp"
/dev_stage="0-7 DAP (days after pollination)"
/lab_host="XL10-Gold"
/clone_lib="HZ"

/note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); pericarp 0-7 DAP (days after pollination). Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To

ORIGIN

Query Match 5.4%; Score 43.4; DB 6; Length 491;
Best Local Similarity 48.2%; Pred. No. 0.35;
Matches 122; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 406 TTCGACCTCGACCGCTCTATTTCAGAGCAATTTGTGCCCAATCGATCTTTTACCGC 465
DB 209 TTCTACTGGGCGCCCTCTCTCGCGAGTCCAAACGCCGCGCGCTCGAGCACCAGCTC 268
QY 466 CCGTAGCTTTTCGACGCGCATCGGCCCTTACAACTCGCTACCGAGTCTGGCGGAGCTGG 525
DB 269 AACGACCGCTCATCGCGGCGGCCATGACCGCCACTCCGCTTCTGGAGGCGCGCC 328
QY 526 GACTTCAATATTTCGCTGTTCTTCCAAACCGCGGCTGATTTACCCGTTACATGACGCTG 585
DB 329 GACGCTCTCGTCTTCAACTCTCTTACCTCTGCTGATGACCGGCGCAACAGATCCAGATCCTG 388
QY 586 ATTTCCGAATACAACGACATGACCGGCTTCAGCATGACGCGGAGCTGATTAAGGTTTC 645
DB 389 AGGGGCGCGCAACGACATGAGCAAGGACATCGTGAGATGGGGCGGAGGCGGTAC 448
QY 646 AGAAAAACGGCTGC 658
DB 449 AGGCTGGTGTCTGC 461

RESULT 10

BI960016

LOCUS
DEFINITION
(normal) Hordeum vulgare subsp. vulgare cDNA clone HVSMEN0022N21f,
mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Hordeum vulgare subsp. vulgare
Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.

1 (bases 1 to 658)
Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Chin, A., Begum, D.,
Frisch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T.,
Simmons, J., Oates, R. and Main, D.

Development of a genetically and physically anchored EST resource
for barley genomics: Morex rachis cDNA library
Unpublished (2001)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 546
Seq primer: AATTAACCTCACTAAAGG
High quality sequence stop: 641.
Location/Qualifiers

JOURNAL

COMMENT

FEATURES
source

1. 658
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMEN0022N21f"
/tissue_type="Rachis"
/lab_host="TJG121"
/clone_lib="Hordeum vulgare rachis EST library HVCDNA0015"

excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 900 bp"

(normal)"
/note=Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Plants were grown at Washington State University, Pullman, WA in a greenhouse, the rachises were excised and frozen in liquid nitrogen (Kleinhofs lab). In the TJ Close lab at the University of California, Riverside total RNA was prepared, poly(A) was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

ORIGIN

Query Match 5.4%; Score 43.4; DB 4; Length 658;
Best Local Similarity 48.2%; Pred. No. 0.36;
Matches 122; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 406 TTCGACCTCGACCGCTCTCTATTTCGAGACGAATTTGTGCCCAATCGATCTTTTACCGC 465
Db 193 TTCTACTGGCGCGCTCTCCCGAGTCCACGCCGACGCGGTGAGCACCAGCTC 252

QY 466 CGTGAGCTTTTCGACGGCATCGCCCTTTACACCTCGCTACCGCTCTGGCGGACTGG 525
Db 253 AACGACCGCTCATCGGGCGGCCCATGGACGCCACTCCGCTTCTGGAAGGCGCC 312

QY 526 GACTTCAATATTCGTGCTTCTCCACCGCGCTGATTACCGCTACATGACGCTGTG 585
Db 313 GAGTCTCTGCTTCAACTCTCTGTTGGATGACCGGCGACAAGATCCAGATCTCTG 372

QY 586 ATTTCCGAATACACGACATACCGCTTTCACATGAGCGAGGACGACTGATAAGAGTTC 645
Db 373 AGGGCGCGCACACGACATGACGACATCTGTGGAGATGGGGCGGAGGCGGTAC 432

QY 646 AGAAACCGCTGC 658
Db 433 AGGCTGGTCTGC 445

RESULT 11
LOCUS B1959554
DEFINITION HVSMEN020C17f Hordeum vulgare rachis EST library HVCDNA0015 (normal) Hordeum vulgare subsp. vulgare cDNA clone HVSMEN020C17f, mRNA sequence.
ACCESSION B1959554
VERSION B1959554.1 GI:16310809
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 669)
AUTHORS Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Chin, A., Begum, D., Frisch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Oates, R. and Main, D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex rachis cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA

Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 418
Seq primer: AATTAACCTCACTAAAGGG
High quality sequence stop: 583.
Location/Qualifiers
1. 669

FEATURES

source
1. 669
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMEN020C17f"
/tissue_type="Rachis"
/lab_host="TJG121"
/clone_lib="Hordeum vulgare rachis EST library HVCDNA0015 (normal)"
/note=Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Plants were grown at Washington State University, Pullman, WA in a greenhouse, the rachises were excised and frozen in liquid nitrogen (Kleinhofs lab). In the TJ Close lab at the University of California, Riverside total RNA was prepared, poly(A) was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

ORIGIN

Query Match 5.4%; Score 43.4; DB 4; Length 669;
Best Local Similarity 48.2%; Pred. No. 0.36;
Matches 122; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 406 TTCGACCTCGACCGCTCTCTATTTCGAGACGAATTTGTGCCCAATCGATCTTTTACCGC 465
Db 130 TTCTACTGGCGCGCTCTCTCCGAGTCCACGCCGACGCGGTGAGCACCAGCTC 189

QY 466 CGTGAGCTTTTCGACGGCATCGCCCTTTACAACTCGCTACCGAGTCTGGCGGACTGG 525
Db 190 AACGACCGCTCATCGGGCGGCCCATGACCCCACTCCGCTTCTGGAAGGCGCC 249

QY 526 GACTTCAATATTCGTGCTTCTCCAAACCGCGCTGATTACCCGCTACATGACGCTGTG 585
Db 250 GAGTCTCTGCTTCAACTCTCTCTGTGGATGACCGGCGACAAGATCCAGATCTCTG 309

QY 586 ATTTCCGAATACACGACATACCGCTTTCAGCATGAGCGGACCTGATAAGAGTTC 645
Db 310 AGGGCGCGCACACGACATGACGAAGACATCTGTGGAGATGGGGCGGAGGCGGTAC 369

QY 646 AGAAACCGCTGC 658
Db 370 AGGCTGGTCTGC 382

RESULT 12
AJ274182

LOCUS AJ274182 608 bp mRNA linear EST 15-NOV-1999
 DEFINITION AJ274182 Metarhizium anisopliae ARSEF 2575 Metarhizium anisopliae
 cDNA clone Ma#1818, mRNA sequence.
 ACCESSION AJ274182
 VERSION AJ274182.1 GI:64333555
 KEYWORDS EST.
 SOURCE Metarhizium anisopliae
 ORGANISM Metarhizium anisopliae
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic
 Clavicipitaceae; Metarhizium.
 1 (bases 1 to 608)
 REFERENCE Screen, S.E., Mathur, P. and St. Leger, R.J.
 AUTHORS Screen, S.E., Mathur, P. and St. Leger, R.J.
 TITLE EST analysis of the insect pathogenic fungus Metarhizium anisopliae
 JOURNAL Unpublished (1999)
 COMMENT Contact: Screen SE
 Entomology
 University of Maryland
 4112 Plant Sciences Building, College Park, MD 20742, USA.
 Location/Qualifiers
 1. .608
 /organism="Metarhizium anisopliae"
 /mol_type="mRNA"
 /strain="ARSEF 2575"
 /db_xref="taxon:5530"
 /clone="Ma#1818"
 /note="Metarhizium anisopliae ARSEF 2575"
 /vector="Unizap; Metarhizium anisopliae was grown on
 insect cuticle for 24 hours. A cDNA library was
 constructed in the unidirectional Lambda vector, Unizap"

ORIGIN
 Query Match 5.4%; Score 43.2; DB 1; Length 608;
 Best Local Similarity 52.7%; Pred. No. 0.41;
 Matches 118; Conservative 0; Mismatches 103; Indels 3; Gaps 1;
 QY 74 GGATGTCGGGAGAGCTACCGGAGTGAAGTGGTCTTTCGACGGGTTTCGACCG 133
 Db 248 GCACCTTTTACCGAGAGCAATCGATTGGGAATCATATTCGACGAGCGTTGCCCG 307
 QY 134 ATCGGACCTTCGACGCGGACAGTTTCGGCCGGAATCGGCTCGGACTGGTTCGTC 193
 Db 308 ATGGACCCCAAGAAGTCGCAACCGAGCTGGTCAAGGCTTACGCTCTCATCGTCATCCCTCA 367
 QY 194 ACAGCGGGCCGATGAT---GGCCCCCTACGAGCCATGAACCGCGCGTGGCGGTAGCCA 250
 Db 368 AGACTCGCTCGGCAAGCTCGTCTCGGAACCGCTACGTCACGCGCTGCAGTTCGTCA 427
 QY 251 CAGGCGAATGGGTACTTTTTTTAGGCGCGGACGACACCTCTTAC 294
 Db 428 CCGGCAACTTTGTTCATCATCATGATGCGGACTTCTCCCAACCAC 471

RESULT 13
 LOCUS CK201078 802 bp mRNA linear EST 08-DEC-2003
 DEFINITION FGAS009597 Triticum aestivum FGAS: Library 3 Gate 6 Triticum
 aestivum cDNA, mRNA sequence.
 ACCESSION CK201078
 VERSION CK201078.1 GI:39563468
 KEYWORDS EST.
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Pooidae; Triticeae; Triticum.
 1 (bases 1 to 802)
 REFERENCE Allard, P., Crosby, W.L., Danyluk, J., Rudes, P., Frick, M., Gaudet, D.,
 Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Larocche, A.,
 Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,
 Penniket, C., Roach, J.L. and Sarhan, F.
 Functional Genomics of Abiotic Stress in Wheat and Canola Crops
 Unpublished (2003)

LOCUS CK201078 464 bp mRNA linear EST 23-OCT-2002
 DEFINITION HZ52A13r HZ Hordeum vulgare subsp. vulgare cDNA clone HZ52A13
 5-PRIME, mRNA sequence.
 ACCESSION CA025392
 VERSION CA025392.1 GI:24302766
 KEYWORDS EST.
 SOURCE Hordeum vulgare subsp. vulgare
 ORGANISM Hordeum vulgare subsp. vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Pooidae; Triticeae; Hordeum.
 1 (bases 1 to 464)
 REFERENCE Radchuk, V., Zhang, H., Weschke, W., Potokina, E. and Wobus, U.
 AUTHORS Radchuk, V., Zhang, H., Weschke, W., Potokina, E. and Wobus, U.
 TITLE Barley ESTs from developing seeds
 JOURNAL Unpublished (2002)
 COMMENT Contact: Stein Nils
 Molecular Markers Group, Department Genbank.

COMMENT Contact: Wm L Crosby
 Bioinformatics
 University of Saskatchewan, Department of Computer Science
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,
 Saskatchewan, S7N 5A9, Canada
 Tel: 306 966 1769
 Fax: 306 966 2033
 Email: fgas_est@cs.usask.ca
 This sequence is the direct result of the Base calling software
 Phred (default parameters). It is the raw base calls. To aid in the
 identification of the high quality insert the software Lucy
 (default parameters) has been run on this sequence. Lucy identified
 the region [101,763].
 Plate: L3C113 row: C column: 12.
 Location/Qualifiers
 1. .802
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /db_xref="taxon:4565"
 /clone_lib="Triticum aestivum FGAS: Library 3 Gate 6"
 /note="Organ: Root; Vector: pCMV.SPORT6; Root tissue from
 control, cold-acclimated and salt stressed wheat cultivar
 Norstar. 7 mRNA populations were combined before
 constructing the library; 7 day non-acclimated roots, 1,
 23, and 53 days cold-acclimated at 4C, and 30 minutes, 3
 hours and 6 hours treated roots with 200mM NaCl.
 Non-acclimated and cold-acclimated plants were grown in
 vermiculite while salt stressed plant were grown
 hydroponically. First strand synthesis in this library was
 done in the presence of methylated dCTP thereby protecting
 from internal cleavage with NotI."

FEATURES
 source

ORIGIN
 Query Match 5.4%; Score 43; DB 7; Length 802;
 Best Local Similarity 50.2%; Pred. No. 0.49;
 Matches 106; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
 QY 47 CGGTGACGCTGCAACGCTCCCTCGGAAGCATCTCGGAGCAGACCTTACCGGGAAGTGGAG 106
 Db 253 CCGTGACGCGCAGGAGCGGCTGGAGCCATCGCCCGCAACTCTTCGACGCGTGTCA 312
 QY 107 TGGTCTTTGTCGACGGCGGTTGACCGATCGGACCTCGACATCGGGAACAGTTTCGCGC 166
 Db 313 CCTACGAGGATGCCACCGCCCAACAACATCGCGGAGCTCAACCTCATCACCGTCGGC 372
 QY 167 CGGAACCTCGGCTCGGACTGCTGTTACAGCGGCGCGATGATGCGCCCTACGACGCCA 226
 Db 373 AGAAGCTTGGATCCGCTGCTGCTGACGTCGACCCGTCGAGCGGCGCGCGCGCTCT 432
 QY 227 TGAACCGCGGCTCGGCTAGCCACAGCGCA 257
 Db 433 TCCACCTGCGCCACATCTCGACGCGCGGGA 463

RESULT 14
 LOCUS CA025392
 DEFINITION HZ52A13r HZ Hordeum vulgare subsp. vulgare cDNA clone HZ52A13
 5-PRIME, mRNA sequence.
 ACCESSION CA025392
 VERSION CA025392.1 GI:24302766
 KEYWORDS EST.
 SOURCE Hordeum vulgare subsp. vulgare
 ORGANISM Hordeum vulgare subsp. vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Pooidae; Triticeae; Hordeum.
 1 (bases 1 to 464)
 REFERENCE Radchuk, V., Zhang, H., Weschke, W., Potokina, E. and Wobus, U.
 AUTHORS Radchuk, V., Zhang, H., Weschke, W., Potokina, E. and Wobus, U.
 TITLE Barley ESTs from developing seeds
 JOURNAL Unpublished (2002)
 COMMENT Contact: Stein Nils
 Molecular Markers Group, Department Genbank.

Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 464 Std Error: 0.00
Plate: 52 row: A column: 13
Seq primer: M13rev.

FEATURES
source

Location/Qualifiers
1. .464
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="barke"
/sub_species="vulgare"
/db_xref="GABI:278814"
/db_xref="taxon:112509"
/clone="HZ52A13"
/tissue_type="pericarp"
/dev_stages="0-7 DAP (days after pollination)"
/lab_host="XL10-Gold"
/clone_lib="HZ"

/note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); pericarp 0-7 DAP(days after pollination). Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used blue/white selection for recombinants is not 100% reliable.Average insert size is 900 bp"

ORIGIN

Query Match 5.3%; Score 42.4; DB 6; Length 464;
Best Local Similarity 47.8%; Pred. No. 0.67;
Matches 121; Conservative 0; Mismatches 132; Indels 0; Gaps 0;
QY 406 TTCGACCTCGACCGCTCTATTTCGAGAGGATTTGTGCCACCAATCGATCTTTTACCGC 465
DB 209 TTCTACTGGCGCGCTCTCTCCGAGTCCAAACGCCGACGCGGTGAGCACCAGCTC 268
QY 466 CGTGAGCTTTTCGAGCGCATCGCCCTTACAACTCGCGCTACCGAGTCTGGCGGACTGG 525
DB 269 AACGACCGCTCATCGCGCGCGCCCATGGACCGCCACTCCGCTTCTGGAGGGCGCC 328
QY 526 GACTTCAATATTCGTGCTTCTCCAAACCGCGGCTGATTACCCGCTCATGAGACGCTCGT 585
DB 329 GACGTCCTCGTCTTCAACTCTTACCTCTGTGTGATGACCGGCGACAAGATCCAGATCCTG 388
QY 586 ATTTCCGAATACAGCATGACCGCTTCACATGAGCGAGCGGAGCTGATAAAGATTTC 645
DB 389 AGGGGCGCGCAACGACATGAGCAAGACATCGTGGAGATGNGGCGGAGGCGGTAC 448
QY 646 AGAAAACGGCTGC 658
DB 449 AGGCTGTGCTGC 461

RESULT 15
CA023059
LOCUS HZ45B23: HZ Hordeum vulgare subsp. vulgare cDNA clone HZ45B23
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION CA023059
VERSION CA023059.1 GI:24300433
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 553)
AUTHORS Radchuk,V., Zhang,H., Weschke,W., Potokina,E. and Wobus,U.
TITLE Barley ESTs from developing seeds

JOURNAL
COMMENT

Unpublished (2002)
Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 553 Std Error: 0.00
Plate: 45 row: B column: 23
Seq primer: M13rev.

FEATURES

source

Location/Qualifiers
1. 553
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="barke"
/sub_species="vulgare"
/db_xref="GABI:276467"
/db_xref="taxon:112509"
/clone="HZ45B23"
/tissue_type="pericarp"
/dev_stages="0-7 DAP (days after pollination)"
/lab_host="XL10-Gold"
/clone_lib="HZ"

/note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); pericarp 0-7 DAP(days after pollination). Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable.Average insert size is 900 bp"

ORIGIN

Query Match 5.3%; Score 42.4; DB 6; Length 553;
Best Local Similarity 47.8%; Pred. No. 0.68;
Matches 121; Conservative 0; Mismatches 132; Indels 0; Gaps 0;
QY 406 TTCGACCTCGACCGCTCTATTTCGAGAGGATTTGTGCCACCAATCGATCTTTTACCGC 465
DB 301 TTCTACTGGCGCGCTCTCTCCGAGTCCAAACGCCGACGCGGTGAGCACCAGCTC 360
QY 466 CGTGAGCTTTTCGAGCGCATCGCCCTTACAACTCGCGCTACCGAGTCTGGCGGACTGG 525
DB 361 AACGACCGCTCATCGCGCGCGCCCATGGACCGCCACTCCGCTTCTGGAGGGCGCC 420
QY 526 GACTTCAATATTCGTGCTTCTTCCAAACCGCGGCTGATTACCCGCTACATGAGCTCGTG 585
DB 421 GACGTCCTCGTCTTCAACTCTTACCTCTGTGTGATGACCGGCGACAAGATCCAGATCCTG 480
QY 586 ATTTCCGAATACAGCATGACCGGCTTTCAGCATGAGCGGAGGAGCTGATAAAGATTTC 645
DB 481 AGGGGCGCGCAACGACATGAGCAAGACATCGTGGAGATGNGGCGGAGGCGGTAC 540
QY 646 AGAAAACGGCTGC 658
DB 541 AGGCTGTGCTGC 553

Search completed: March 6, 2005, 22:01:40
Job time : 3394 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2005, 19:42:08 ; Search time 185 Seconds
(without alignments)
7084.636 Million cell updates/sec

Title: US-10-805-311-23

Perfect score: 801

Sequence: 1 atgactgcgcagtggtctc.....aacgaagcgcagaccgtag 801

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	346	43.2	32155	US-08-311-731A-1	Sequence 1, Appli
2	346	43.2	4403765	US-09-103-840A-2	Sequence 2, Appli
3	346	43.2	4411529	US-09-103-840A-1	Sequence 1, Appli
4	58.4	7.3	912	US-09-252-991A-12584	Sequence 12584, A
5	56.4	7.0	906	US-09-252-991A-16525	Sequence 16525, A
C 6	55.6	6.9	1060	US-09-072-596-306	Sequence 306, App
7	55.6	6.9	1060	US-09-072-967-311	Sequence 311, App
8	47.8	6.0	1022	US-09-072-596-325	Sequence 325, App
9	47.8	6.0	1022	US-09-072-967-330	Sequence 330, App
10	43.8	5.5	1332	US-09-902-540-4672	Sequence 4672, App
11	43.8	5.5	24754	US-09-902-540-1230	Sequence 1230, App
12	43.6	5.4	36412	US-08-311-731A-132	Sequence 132, App
13	42.8	5.3	5859	US-08-312-387B-1	Sequence 1, Appli
14	42.8	5.3	5859	US-08-312-387B-7	Sequence 7, Appli
15	42.8	5.3	5859	US-08-683-426-1	Sequence 1, Appli
16	42.8	5.3	5859	US-08-683-426-7	Sequence 7, Appli
17	42.8	5.3	5859	US-08-683-458-1	Sequence 1, Appli
18	42.8	5.3	5859	US-08-683-458-7	Sequence 7, Appli
19	42.8	5.3	5859	US-08-878-360-1	Sequence 1, Appli
20	42.8	5.3	5859	US-08-878-360-7	Sequence 7, Appli
21	42.8	5.3	5859	US-08-478-140B-1	Sequence 1, Appli
22	42.8	5.3	5859	US-09-333-412-1	GENERAL INFORMA
23	42.8	5.3	5859	US-09-333-412-7	GENERAL INFORMA
24	42.8	5.3	5859	US-09-338-943-1	Sequence 1, Appli
25	42.8	5.3	5859	US-10-007-267A-1	Sequence 1, Appli
26	42.8	5.3	5859	US-10-007-267A-7	Sequence 7, Appli
27	42.6	5.3	1431	US-09-489-039A-1316	Sequence 1316, App

C 28	42	5.2	2730	4	US-09-902-540-5705	Sequence 5705, Ap
29	42	5.2	50725	4	US-09-902-540-1271	Sequence 1271, Ap
30	41.8	5.2	1131	4	US-09-902-540-6687	Sequence 6687, Ap
C 31	41.8	5.2	9165	4	US-09-902-540-1050	Sequence 1050, Ap
32	41.4	5.2	1521	4	US-09-252-991A-11805	Sequence 11805, A
C 33	41.4	5.2	2505	4	US-09-252-991A-11574	Sequence 11574, A
C 34	41.2	5.1	1443	4	US-09-252-991A-12762	Sequence 12762, A
35	40.8	5.1	320	3	US-09-060-756-467	Sequence 467, App
36	40.8	5.1	320	4	US-09-670-314-467	Sequence 467, App
37	40.8	5.1	1836	4	US-09-266-965-52	Sequence 52, Appl
38	40.8	5.1	53500	4	US-09-266-965-76	Sequence 76, Appl
39	40.4	5.0	2811	4	US-09-266-965-28	Sequence 28, Appl
40	40.4	5.0	18034	4	US-09-266-965-75	Sequence 75, Appl
41	39.2	4.9	786	4	US-09-894-844-51	Sequence 51, Appl
42	39.2	4.9	2004	4	US-09-252-991A-11865	Sequence 11865, A
C 43	39.2	4.9	12732	3	US-09-060-756-1	Sequence 1, Appli
C 44	39.2	4.9	12732	4	US-09-670-314-1	Sequence 1, Appli
C 45	39.2	4.9	4403765	3	US-09-103-840A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-08-311-731A-1/c
; Sequence 1, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/311.731A
; APPLICATION NUMBER: US/08/311.731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32155 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: MYCOBACTERIUM TUBERCULOSIS
; US-08-311-731A-1

Query Match 43.2%; Score 346; DB 4; Length 32155;
Best Local Similarity 71.0%; Pred. No. 5.7e-95;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

```

QY 1 ATGACTGCGCGAGTGTCTCGATAATATCCCTACCTTCAATGACAGCGGTGACGCTGCAA 60
Db 15076 ATGCGCGACCAATGTTTTCGATCATATCCACCCTTTGACAGTGGCTGCGGTATGCT 15017
QY 61 GCCTGCTCGGAAGCATCGTGGGCGAGACCTACCGGGAAGTGGAGTGGTCTTGTTCGAC 120
Db 15016 GCCTGCTCGAGCATCGCCCGTCAGACCTGCGGTGACTTCGAGCTGGTACTGGTCGAC 14957
QY 121 GCGGTTGACCGCATCGGACCTCGACATCGGGAACAGTTTCGCGCCCGGAACCTCGGCTCG 180
Db 14956 GCGGCTCGAGCGAGAAACCTCGACATCGCAACATTTTCGCCCCCAACCTCGCGCAG 14897
QY 181 CGACTGCTGCTTCACAGCGGGCCGATGATGCCCCCTACGAGCGCATGAACCGGGCGTGC 240
Db 14896 CGGTTGATCATTCATCGCGACACCGAGCGGTCTACGAGCGCATGAACCGGGCGGTG 14837
QY 241 GCGTAGGACAGGCGAATGGGTACTTTTATAGCGCGCGACGACACCTCTACGAACCA 300
Db 14836 GACCTGGCCACCGGAACGTGGTTGCTCTTCTGGCGCGGACGACGCTGTACGAGGCT 14777
QY 301 ACCAGTTGGCCAGGTAGCGCTTTTCTCGGCGACCATGCGGCAAGCCATCTTGTCTAT 360
Db 14776 GACACCTGGCGGGTGGCGCTTTCATTTGGCGAAACAGAGCCCGAGCGATCTGTATAT 14717
QY 361 GCGGATGTTGATCGGTTTCGACGAAAGCGGCATCGCGACCTTTCGACCTCGACCGC 420
Db 14716 GCGGACGTGATCATCGCTTCAACCAATTTTCGCTGGGGTGGCGCTTCGACCTCGACCGT 14657
QY 421 CTCCTATTGAGACGAATTTGTGCAACCAATGATCTTTTACCGCGGTGAGCTTTTCGAC 480
Db 14656 CTGTTGTTCAAGCGCAACATCTGCCATCAGGGATCTTCTACCGCGGAGCTCTTCGCG 14597
QY 481 GGCATCGGCGCTTACAACTCGGCTTACGAGTCTGGGCGGAGCTGGGAATTCATATTCGC 540
Db 14596 ACCATCGGTCCCTACAACTCGCTTACCGGGTCTTGGCGGAGCTTCAATATTCGC 14537
QY 541 TGCTTCTCCACCGCGCTGATTTACCGCTACATGAGCGCTGATTTCCGATATACAAAC 600
Db 14536 TGCTTTTCCACCGCGCTGCTACCGCTTACCGCTTACATGAGCGGTGCTTCAATATTCGC 14477
QY 601 GACATGACCGGCTTCAGCATGAGGCGGAGTGTATAAGAGTTTCAGAAAACGGCTGCCA 660
Db 14476 GAATTCGGGGCTCAGCAATACGATCG---TCGACAGGAGTTTTCGAGCGGCTGCGG 14420
QY 661 ATGTAC 666
Db 14419 ATGTCC 14414

```

```

RESULT 2
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-2007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: "n" bases at various positions throughout the sequence

```

```

; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
Query Match 43.2%; Score 346; DB 3; Length 4403765;
Best Local Similarity 71.0%; Pred. No. 5.6e-94;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;
QY 1 ATGACTGCGCGAGTGTCTCGATAATATCCCTACCTTCAATGACAGGCGTACGCTGCAA 60
Db 3303798 ATGCGCGACCAATGTTTTCGATCATATCCACCCTTTGACAGTGGCTGCGGTATGCT 3303857
QY 61 GCCTGCTCGGAAGCATCGTGGGCGAGACCTACCGGGAAGTGGAGTGGTCTTGTTCGAC 120
Db 3303858 GCCTGCTCGACAGCATCGCCCGTCAGACCTGCGGTGACTTCGAGCTGGTACTGGTCGAC 3303917
QY 121 GCGGTTGACCGCATCGGACCTCGACATCGGGAACAGTTTCGCGCCCGGAACCTCGGCTCG 180
Db 3303918 GCGGCTCGACGAGAAACCTCGACATCGCAACATTTTCGCCCCCAACCTCGCGCAG 3303977
QY 181 CGACTGCTGCTTCACAGCGGGCCGATGATGCCCCCTACGAGCGCATGAACCGGGCGTGC 240
Db 3303978 CGGTTGATCATTCATCGCGACACCGAGCGGTCTACGAGCGCATGAACCGGGCGGTG 3304037
QY 241 GCGTAGGACAGGCGAATGGGTACTTTTATAGCGCGCGACGACACCTCTACGAACCA 300
Db 3304038 GACCTGGCCACCGGAACGTGGTTGCTCTTCTGGCGCGGACGACGCTGTACGAGGCT 3304097
QY 301 ACCAGTTGGCCAGGTAGCGCTTTTCTCGGCGACCATGCGGCAAGCCATCTTGTCTAT 360
Db 3304098 GACACCTGGCGGGTGGCGCTTTCATTTGGCGAAACAGAGCCCGAGCTCTGTATAT 3304157
QY 361 GCGGATGTTGATCGGTTTCGACGAAAGCGGCATCGCGACCTTTCGACCTCGACCGC 420
Db 3304158 GCGGACGTGATCATCGCTTCAACCAATTTTCGCTGGGGTGGCGCTTCGACCTCGACCGT 3304217
QY 421 CTCCTATTGAGACGAATTTGTGCAACCAATGATCTTTTACCGCGGTGAGCTTTTCGAC 480
Db 3304218 CTGTTGTTCAAGCGCAACATCTGCCATCAGCGGATCTTACCGCGCGGAGCTCTTCGCG 3304277
QY 481 GGCATCGGCGCTTACAACTCGGCTTACGAGTCTGGGCGGAGCTGGGACTTCAATATTCGC 540
Db 3304278 ACCATCGGTCCCTACAACTCGCTTACCGGTCTTGGCGGAGCTTGGGACTTCAATATTCGC 3304337
QY 541 TGCTTCTCCACCGCGCTGATTTACCGCTACATGAGCGCTGATTTCCGATATACAAAC 600
Db 3304338 TGCTTTTCCACCGCGCTGCTACCGCTTACATGAGCGGTGCTTGGAGCTACAAAC 3304397
QY 601 GACATGACCGGCTTCAGCATGAGGCGGAGTGTATAAGAGTTTCAGAAAACGGCTGCCA 660
Db 3304398 GAATTCGGGGCTCAGCAATACGATCG---TCGACAGGAGTTTTCGAGCGGCTGCGG 3304454
QY 661 ATGTAC 666
Db 3304455 ATGTCC 3304460

```

```

RESULT 3
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-2007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1

```

LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 43.2%; Score 346; DB 3; Length 4411529;
Best Local Similarity 71.0%; Pred. No. 5.6e-94;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

QY 1 ATGACTGGCCAGTGTTCGATATATCCCTACCTTCAATGACGCGGTGACGCTGCAA 60
DB 3309524 ATGGCCGACCACTGTTTCGATCATATCCCACTTGAACGTGGCTGCGGTATTCCT 3309583
QY 61 GCTTCCTCGAAGCATCGTCGGGAGAGCTTACCGGGAAGTGGAAAGTGGTCTCTTCGAC 120
DB 3309584 GCTTCCTCGAAGCATCGTCGGGAGAGCTTACCGGGAAGTGGAAAGTGGTCTCTTCGAC 3309643
QY 121 GCGGTTCGACGATCGGACCTCGACATCGGAAACAGTTTCGCGCCGGAACCTCGGCTCG 180
DB 3309644 GCGGTTCGACGATCGGACCTCGACATCGGAAACAGTTTCGCGCCGGAACCTCGGCTCG 3309703
QY 181 CGACTGTCTGTCACAGCGGCGCCGATGATGSCCCCTACGAGCCATGAACCGGCGCTC 240
DB 3309704 CGTTGATCATTCATCGCGACACCGAGGCGCTCTACGAGCCATGAACCGGCGCTC 3309763
QY 241 GCGGTAGCCACAGGCGAATGGGTACTTTTATAGCGCGCGACGACACCTCTACGAACCA 300
DB 3309764 GACCTGGCCACCGGAACGTGGTGTCTTTCTGGGCGGAGACAGCTGTACGAGCT 3309823
QY 301 ACCAGTTGGCCAGGTAGCGCTTTCTCGGCGACCAATGCGGCAAGCCATCTTGTCTAT 360
DB 3309824 GACACCTGGGCGGCTGGCGCCCTTCATTTGGCGAAACAGAGCCGAGCATCTGGTATAT 3309883
QY 361 GCGGATGTTGATCGGTTCGAGGAAGCGGCATCGCGACCTTCGACCTCGACCGC 420
DB 3309884 GCGGATGTTGATCGGTTCGAGGAAGCGGCATCGCGACCTTCGACCTCGACCGC 3309943
QY 421 CTCCTATTGAGACGAATTTGTGCAACCAATGATCTTTTACCGCGGTGAGCTTTTCGAC 480
DB 3309944 CTGTTGTTCAAGCGCAACATCTGCATCAGGGGATCTTACCGCGGAGCTTTTCGCG 3310003
QY 481 GGCATGCGCTTACAACTCGGCTACCGAGTCTGGGCGGAGCTGGGACTTCAATATTGCG 540
DB 3310004 ACCATCGCTTCCATACAACTCGGCTACCGGCTTGGGCGGAGCTTCAATATTGCG 3310063
QY 541 TGCTTCTCAACCCGCGCTGATACCGCTACATGAGGCTGCTGATTTCCGATACAACT 600
DB 3310064 TGCTTCTCAACCCGCGCTGATACCGCTACATGAGGCTGCTGATTTCCGATACAACT 3310123
QY 601 GACATGACCGGCTTCAGCATGAGGCGGAGCTGATAAGAGTTTCAGAAAACGGCTGCGA 660
DB 3310124 GAATTCGCGGGCTCAGCAATACGATCG---TCGACAGGAGTTTTCAGAGGCTGCGG 3310180
QY 661 ATGTAC 666
DB 3310181 ATGTCC 3310186

RESULT 4
US-09-252-991A-12584
Sequence 12584, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 12584
LENGTH: 912
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12584

Query Match 7.3%; Score 58.4; DB 4; Length 912;
Best Local Similarity 49.8%; Pred. No. 1.2e-07;
Matches 208; Conservative 0; Mismatches 201; Indels 9; Gaps 2;

QY 7 GCGCAGGTTCCTCGATAATTCCTTACCTTCAATGACGCGGTGACGCTGCAAGCCTGCG 66
DB 58 GCGCAGGTTCCTCGATAATTCCTTACCTTCAATGACGCGGTGACGCTGCAAGCCTGCG 117
QY 67 CTCGGAAGCATCGTCGGGAGAGCTTACCGGGAAGTGGAAAGTGGTCTTGTTCGACGCGGT 126
DB 118 CTGCGCAGCATCTACGAGCAGGACTACCCGAAATTCGAAAGTGATCATCTGTCGACGCGT 177
QY 127 TCGACCGATCGGACCTTCGACATCGCAACAGTTTCGCGCCGGAACCTCGGCTCGGACTG 186
DB 178 TCCACCGCAACAGCTAGCCATCTGAGCAGTTGCAAGAGTCCAGGTTTCCAGC-- 235
QY 187 GTCTTCACAGCGGCGCGATGATGGCCCTTACGACGCTATGAACCGCGGCTGCGGCTA 246
DB 236 ---TCTACCGCAGCAGAACGAGGCGTCAAGCGGCTGAACTTCGGTCTGCAACAC 291
QY 247 GCGACGCGCATGGGTACTTTTATAGGCGCCGACGACCTTACGAAACCAACACG 306
DB 292 GCGCGCGGACTAGTAGCCACCGGACCTGAGCAGCA---TCATGCTCGCGACTCG 348
QY 307 TTGCGCCAGGTAGCGCTTTCTCGGACCATCGGCAAGCCATCTTGTCTATGGCGAT 366
DB 349 TTGAGCGTGGCGCGGCTACTCGACGACCGGAGTGGGCTGGGCGGCTG 408
QY 367 GTTGTGATGCGTTCGACGAAAGCGGCGATGCGGACCTTTTCGACCTCGACCGGCTCC 424
DB 409 GTGATCTACATCGACAGCGAGGCGCAGGAAACCAAGCGGAGCAACCGCATCC 466

RESULT 5
US-09-252-991A-16525
Sequence 16525, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 16525
LENGTH: 906
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16525

Query Match 7.0%; Score 56.4; DB 4; Length 906;
Best Local Similarity 56.5%; Pred. No. 4.9e-07;
Matches 105; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 9 GCCAGTGTTCGATAATTCCTTACCTTCAATGACGCGGTGACGCTGCAAGCCTGCT 68
DB 48 GCCATTTGTTTCCGTGATCATCTGCTTCTATACCAAGGAGTACATAGAGCGAGAT 107
QY 69 CGGAAGCATCGTCGGGAGGAGCTTACCGGGAAGTGGTCTTGTTCGACGCGGCTTC 128

Db 108 TGCCAGCGTGGCTGCGCAGACATACCGCAACGTGGAACTGCTGGTGGAGCATGGTTC 167
Qy 129 GACCGATCGGACCTCGACATCGGAACAGTTTCGCCCGGAACCTCGGCTCGGACTGGT 188
Db 168 CAGCGATCGCAGCGTCGAGCTTTTTCGCGCGGCTCCAGGCGAAGTAGTCGACCTCAG 227
Qy 189 GGTCA 194
Db 228 GGTCA 233

RESULT 6
US-09-072-596-306/c
; Sequence 306 Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072.596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 306:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-09-072-596-306

Query Match 6.9%; Score 55.6; DB 3; Length 1060;
Best Local Similarity 50.2%; Pred. No. 9.2e-07;
Matches 121; Conservative 22; Mismatches 95; Indels 3; Gaps 3;
Qy 338 ATGCGGCAAGCATCTTGTCTATGCGGATGTTGTGATCGGTTTCGACGAAAGCCGGCATG 397
Db 744 ANNMGACCCNAKATTTGNAMWGGGAAYNTRWTMTGSGTTC-ACCAATTTCGSGGGG 686
Qy 398 CCGGACCTTCGACCTCGACCGCCTCTATTGAGCAAAATTTGCGCCACCAATCGATCT 457
Db 685 GGGNCCTTTGTCMCTTGACCTTTTGTGTTTCNAGSAAATTS-KCCATNSNGGWWTT 627
Qy 458 TTTACCGCGTGAAGCTTTTTCGACGCGCATCGGCCCTTACAACTCGCGCTACCGAGTCTGGG 517

Db 626 TYACNCCCGGATTTTGGGCAMCAITGGTTTCCAMCMCTTCGKAAACNGGTTCTKTCG 567
Qy 518 CGGACTGGGACTTCAATATTCGCTGCTCTCCACACCGCGCGCTGATTACCGCTACATGG 577
Db 566 CCGANTGGGANTTCAATATTTGGNTGTTTTCNAAACCARG-GTTGTTNACCSGNAACATGN 508
Qy 578 A 578
Db 507 A 507

RESULT 7
US-09-072-967-311/c
; Sequence 311 Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072.967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 311:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-09-072-967-311

Query Match 6.9%; Score 55.6; DB 4; Length 1060;
Best Local Similarity 50.2%; Pred. No. 9.2e-07;
Matches 121; Conservative 22; Mismatches 95; Indels 3; Gaps 3;
Qy 338 ATGCGGCAAGCATCTTGTCTATGCGGATGTTGTGATCGGTTTCGACGAAAGCCGGCATG 397
Db 744 ANNMGACCCNAKATTTGNAMWGGGAAYNTRWTMTGSGTTC-ACCAATTTCGSGGGG 686
Qy 398 CCGGACCTTCGACCTCGACCGCCTCTATTGAGCAAAATTTGCGCCACCAATCGATCT 457
Db 685 GGGNCCTTTGTCMCTTGACCTTTTGTGTTTCNAGSAAATTS-KCCATNSNGGWWTT 627
Qy 458 TTTACCGCGTGAAGCTTTTTCGACGCGCATCGGCCCTTACAACTCGCGCTACCGAGTCTGGG 517

Db 626 TACNCCGGGATTTTGGGCMACATCGTTCCCMACAMCTTCGGAACNGGTTCKTGC 567
Qy 518 CGGACTGGGACTTCAATATTCCTCTCCAAACCGGCGCTGATTACCCGCTACATGG 577
Db 566 CCGANTGGGANTTCAATATTCGNTGTTTTTTCNAACCG- GTTGTNACCSGNAACATGN 508
Qy 578 A 578
Db 507 A 507

RESULT 8
US-09-072-596-325
; Sequence 325, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 325:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1022 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-09-072-596-325

Query Match 6.0%; Score 47.8; DB 3; Length 1022;
Best Local Similarity 49.1%; Pred. NO. 0.00022;
Matches 82; Conservative 20; Mismatches 63; Indels 2; Gaps 1;
Qy 1 ATGACTGGCCAGTGTCTCGATAATTATCCCTACCTTCAATGCGGCTGACCTGCAA 60
Db 405 WTGGCCGCACCAWNNTTTCYCRATCWYCCCCACCTTAACTTGKTYGCSGTATTCCT 464
Qy 61 GCCTGCTTCGGAAGCATCGTTCGGGCGAGACCTTACCGGAAGTGGAGTGGTCTTGTTCGAC 120
Db 465 KCCTGCCTCRACAGCMYCNCCCKTCAACCTGCGGTGACTCCAAGTGT--CTGGYCGAA 522
Qy 121 GCGGTTTCGACCGATCGGACCCCTCGACATCGGAACAGTTTTCCGCC 167

US-09-072-540-4672
; Sequence 4672, Application US/09902540

Db 523 SGGGGGYTCAMCGGACAAACCCCRANNTTCGCAAAATTTTCNCCCC 569
RESULT 9
US-09-072-967-330
; Sequence 330, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 330:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1022 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-09-072-967-330

Query Match 6.0%; Score 47.8; DB 4; Length 1022;
Best Local Similarity 49.1%; Pred. NO. 0.00022;
Matches 82; Conservative 20; Mismatches 63; Indels 2; Gaps 1;
Qy 1 ATGACTGGCCAGTGTCTCGATAATTATCCCTACCTTCAATGCGGCTGACCTGCAA 60
Db 405 WTGGCCGCACCAWNNTTTCYCRATCWYCCCCACCTTAACTTGKTYGCSGTATTCCT 464
Qy 61 GCCTGCTTCGGAAGCATCGTTCGGGCGAGACCTTACCGGAAGTGGAGTGGTCTTGTTCGAC 120
Db 465 KCCTGCCTCRACAGCMYCNCCCKTCAACCTGCGGTGACTCCAAGTGT--CTGGYCGAA 522
Qy 121 GCGGTTTCGACCGATCGGACCCCTCGACATCGGAACAGTTTTCCGCC 167
Db 523 SGGGGGYTCAMCGGACAAACCCCRANNTTCGCAAAATTTTCNCCCC 569

RESULT 10
US-09-072-540-4672
; Sequence 4672, Application US/09902540

Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 4672
; LENGTH: 1332
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-4672

Query Match 5.5%; Score 43.8; DB 4; Length 1332;
Best Local Similarity 49.0%; Pred. No. 0.0042; Mismatches 122; Indels 0; Gaps 0;
Matches 117; Conservative 0;
QY 53 CGCTGCAAGCCTGCTCGGAAGCATGCTGGGCGAGACCTACCGGGAAGTGGAAAGTGTTCC 112
DB 170 CCTGGAGAGTCTTCGGTGGCATGGGCGGAGCGGAGCGTCAGAGAATCTCCGGCC 229
QY 113 TTGTCGACGGGGTTGACCGGATCGGACCTCGACATCGGAACAGTTTCGCGCCCGGAAC 172
DB 230 TGGGCGAGGGGCCCATCATCGACCCGAGCGGCATCGTCAACCAATGACCACGTCATCC 289
QY 173 TCGGCTCGGACTGCTGCTTACAGCGGGCCGATGATGCGCCCTACGACGCCATGAACC 232
DB 290 GGGGCGCTTCGCGCATCCACGTGTCCTGGGAGATGGCGCTCGCTTCGACCGGAGGTCA 349
QY 233 GCGGCGCTGGGCTAGCCAGCGCAATGGGTACTTTTTTTAGGGCCCGACGACACCCCTC 291
DB 350 TCGGCGAGCGCGCGGAACGACCTAGCGGTGCTCAAGTCAATGCCAAGGAGGCCCTC 408

RESULT 11
US-09-902-540-1230
; Sequence 1230, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1230
; LENGTH: 24754
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1230

Query Match 5.5%; Score 43.8; DB 4; Length 24754;
Best Local Similarity 49.0%; Pred. No. 0.016; Mismatches 122; Indels 0; Gaps 0;
Matches 117; Conservative 0;
QY 53 CGCTGCAAGCCTGCTCGGAAGCATGCTGGGCGAGACCTACCGGGAAGTGGAAAGTGTTCC 112
DB 1780 CCTGGAGAGTCTTCGGTGGCATGGGCGGAGCGGAGCGTCAGAGAATCTCCGGCC 1839
QY 113 TTGTCGACGGGGTTGACCGGATCGGACCTCGACATCGGAACAGTTTCGCGCCCGGAAC 172
DB 1840 TGGGCGAGGGGCCCATCATCGACCCGAGCGGCATCATGTCACCAATGACCACGTCATCC 1899

QY 173 TCGGCTCGGACTGCTGTTCCACAGCGGCGCGGATGATGGCCCTAGACGCCATGAACC 232
DB 1900 GGGGCGCTTCGGCCATCACTGTCATCTGGCGAATGCGGCTCTGTTGACCGGAGGTCA 1959
QY 233 GCGGCGTGGGCTAGCCAGCGGAATGGGTACTTTTTTTAGGGCCCGACGACACCCCTC 291
DB 1960 TCGGCGAGCGCGCGGAACGACCTAGCGGTGCTCAAGTCAATGCCAAGGAGGCCCTC 2018

RESULT 12
US-08-311-731A-132
; Sequence 132, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 132:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 36412 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: MYCOBACTERIUM LEPRAE
US-08-311-731A-132

Query Match 5.4%; Score 43.6; DB 4; Length 36412;
Best Local Similarity 50.5%; Pred. No. 0.022; Mismatches 106; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
QY 150 CGCGAAGAGTTTCCGCGCGGAACCTCGGCTCGGACTGGTCTTCACAGCGGCGCGGATGA 209
DB 12241 CGCTATCTGTCGCGCATGGTTGCAGACCTCGTCCCATCTGCTGACCTGCGCGTGG 12300
QY 210 TGGCCCTTACGACGCCATGAACCGCGCGTGGGCTAGCCACAGCGGAATCGGTACTTTT 269
DB 12301 TGATCGCTACACCGTATGGGCACCCCGTTGCGCGAGCGGCGCATGAATGGAGCGGT 12360
QY 270 TTTAGGCGCCGACGACACCTCTTAGAACCAACACAGTTTGGCCCGAGGTAGCGCTTTCT 329
DB 12361 TCTCGGCAAGGACGACCACTGTATGCTGTGAGACAGTCTGCTACCTAGTCTGCAATTTGT 12420

9 GCCAGTGTCTCGATAATTATCCCTACCTTCAATGCAGCGGTGACGCTGCAAGCCTGCCT 68

Qy	69	CGGAGCATCGTCGGGAGACCTTACCGGGAAGTGGAAAGTGGTCTCTTGTGACACGGCGGTTCC	128
Db	3387	GGCGCGTAGTGGGGCAGACTTGGCGCACTTTGGATATTTTGATTGTCGATGACGGCTC	3446
Qy	129	GACCGATCGGACCCCTCGACATCGGGAACAGTTTTCCGCCCGGAACTCGGTCGCGACTGGT	188
Db	3447	GACGGACGGCAGCGCCGCCATTGCCCGCATTTCCAGAAACAGACGCGCAGGATCAGGAT	3506
Qy	189	CGTTACAGCGGGCCCGATGATGSCCCCTACGACGCCATGAAACGGCGGGCTCGCGTA	246
Db	3507	AAATTTCAATCCCGCAATTTGGGCTTTATCGCCTCTTTAAACATCCGGGCTGACACGAA	3564

RESULT 15

```

US-08-683-426-1
; Sequence 1, Application US/08683426
; Patent No. 5705367
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/683,426
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5859 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Neisseria gonorrhoeae
; STRAIN: F62
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..381
; OTHER INFORMATION: glyc (glycyl tRNA synthetase beta chain)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 445..1491
; OTHER INFORMATION: 1gtA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2342..3262

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2005, 18:38:09 ; Search time 559 Seconds
(without alignments)
8482.492 Million cell updates/sec

Title: US-10-805-311-23
Perfect score: 801
Sequence: 1 atgactgcgcagttcttc.....aacgaagcgcagaaccctag 801

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	801	100.0	801	2 AAT74470	Aat74470 Open read
2	801	100.0	4435	2 AAT74479	Aat74479 Part of t
3	799.4	99.8	801	2 AAT74469	Aat74469 Open read
4	799.4	99.8	7995	2 AAT74478	Aat74478 Complete
5	346	43.2	768	8 ACA38491	Aca38491 Prokaryot
6	346	43.2	828	8 AAT74477	Aat74477 Open read
7	346	43.2	828	8 ACA40754	Aca40754 Prokaryot
8	346	43.2	32155	10 ADB74252	Adb74252 Mycobacte
9	346	43.2	110000	4 AAI9682_33	Continuation (34 o
10	346	43.2	110000	4 AAI9683_32	Continuation (33 o
11	346	43.2	110000	4 AAI9683_33	Continuation (34 o
12	249.8	31.2	110000	4 AAI9682_32	Continuation (33 o
13	60	7.5	900	8 ACA42153	Aca42153 Prokaryot
14	58.4	7.3	912	11 ABD13980	Abd13980 Pseudomon
15	57	7.1	349980	6 ABQ81845	Abq81845 Bifidobac
16	56.4	7.0	906	11 ABD17921	Abd17921 Pseudomon
17	55.6	6.9	1060	2 AAZ19211	Aaz19211 M. tuberc
18	55.6	6.9	1060	2 AAZ19423	Aaz19423 M. tuberc
19	55.4	6.9	1033	8 ACA48731	Aca48731 Prokaryot
20	53.8	6.7	1035	8 ACA51327	Aca51327 Prokaryot

21	50.8	6.3	2000	8 ADA71938	Ada71938 Rice gene
22	47.8	6.0	1022	2 AAZ19230	Aaz19230 M. tuberc
23	47.8	6.0	1022	2 AAZ19442	Aaz19442 M. tuberc
24	45.2	5.6	1827	11 ADM01454	Adm01454 Human cDN
25	45	5.6	37500	12 ADH48029	Adh48029 Clone PS3
26	43.6	5.4	36412	10 ADB74383	Adb74383 Mycobacte
27	42.8	5.3	4725	10 AAD55818	Aad55818 Micromono
28	42.8	5.3	5859	2 AAT14061	Aat14061 N. gonorr
29	42.8	5.3	5859	2 AAT49230	Aat49230 Lipo-Olig
30	42.8	5.3	60196	10 AAD55810	Aad55810 Micromono
31	42.6	5.3	1326	8 ACA35579	Aca35579 Prokaryot
32	42.6	5.3	1431	11 ACH95521	Ach95521 Klebsiell
33	42	5.2	110000	4 AAI9682_37	Continuation (38 o
34	41.4	5.2	1521	11 ABD13201	Abd13201 Pseudomon
35	41.4	5.2	2505	11 ABD12970	Abd12970 Pseudomon
36	41.2	5.1	1443	11 ABD14158	Abd14158 Pseudomon
37	40.8	5.1	320	3 ABQ62968	Abq62968 Mycobacte
38	40.8	5.1	882	8 ACA45716	Aca45716 Prokaryot
39	40.8	5.1	1836	3 AAC55818	Aac55818 S. lavend
40	40.8	5.1	1836	10 ADE10237	Adel0237 S. lavend
41	40.8	5.1	53500	3 AAC55842	Aac55842 Complete
42	40.8	5.1	53500	10 ADE10261	Adel0261 S. lavend
43	40.8	5.1	110000	4 AAI9682_01	Continuation (2 of
44	40.8	5.1	110000	4 AAI9683_01	Continuation (2 of
45	40.6	5.1	1185	8 ADA69780	Ada69780 Rice gene

ALIGNMENTS

RESULT 1
AAT74470
ID AAT74470 standard; DNA; 801 BP.
XX
AC AAT74470;
DT 17-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 11-MAR-1998 (first entry)
XX
DE Open reading frame F from the GS region of M. paratuberculosis.
XX
KW GS; pathogenesis island; pathogenic protein; mycobacterial disease;
KW cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
KW vaccine; inflammatory disease; sarcoidosis; Johne's disease; ss.
XX
OS Mycobacterium avium subsp. paratuberculosis.
XX
FH Key Location/Qualifiers
CDS 1..801
FT /*tag= a

PN W09723624-A2.
XX
PD 03-JUL-1997.
XX
PF 23-DEC-1996; 96WO-GB003221.
XX
PR 21-DEC-1995; 95GB-00026178.
XX
PA (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.
XX
PI Hermontaylor J, Doran T, Millar D, Tizard M, Loughlin M, Sumar N;
PI Ford J;
XX
DR WPI; 1997-351061/32.
XX
DR P-PSDB; AAW21775.
XX
PT New isolated pathogenicity island from mycobacteria - used to develop
PT products for detection, diagnosis, prevention and treatment of
PT mycobacteria infections.
XX
PS Claim 6; Page 50; 62pp; English.

The present sequence represents one of eight open reading frames (ORFs) of a novel polynucleotide sequence designated "GS". GS is a pathogenicity island of 8 kb of DNA comprising a core region of 5.75 kb with multiple ORFs and an adjacent transmissible element of 2.5 kb. The ORFs, and also the transmissible element, encode proteins which may be linked to pathogenicity, such as providing receptors for cellular recognition. GS was discovered and characterised using differential DNA analysis technology. It is found within *Mycobacterium paratuberculosis* and it has also been identified in *Mycobacterium avium* subspecies *silvaticum*. These pathogenic mycobacteria cause chronic inflammation of the intestine and Crohn's disease in humans. The protein products of the ORFs of GS can be used for detecting mycobacteria or for diagnosing, treating or preventing mycobacterial disease. In particular they can be used as vaccines for inflammatory diseases such as Crohn's disease or sarcoidosis in humans or John's disease in animals. (Updated on 25-MAR-2003 to correct PI field.) (Updated on 17-OCT-2003 to standardise OS field)

Query Match	100.0%	Score 801;	DB 2;	Length 801;
Best Local Similarity	100.0%;	Pred. No. 6.2e-222;		
Matches 801; Conservative 0;	Mismatch 0;	Indels 0;	Gaps 0;	
QY	1	ATGACTGCGCCAGTGTCTCGATATATATCCCTACCTTCAATGCAGCGGTGACGCTGCAA	60	
DB	1	ATGACTGCGCCAGTGTCTCGATATATATCCCTACCTTCAATGCAGCGGTGACGCTGCAA	60	
QY	61	GCTGCTCTCGGAAGCATCGTGGGCGAGACCTACCGGGAAGTGAAGTGTCTTGTGCGAC	120	
DB	61	GCTGCTCTCGGAAGCATCGTGGGCGAGACCTACCGGGAAGTGAAGTGTCTTGTGCGAC	120	
QY	121	GGCGGTTGCAACCGATCGACCCCTCGACATCGGGAACAGTTTCGCGCCGGAACCTGGCTCG	180	
DB	121	GGCGGTTGCAACCGATCGACCCCTCGACATCGGGAACAGTTTCGCGCCGGAACCTGGCTCG	180	
QY	181	CGACTGCTCGTTTACAGCGGGGCCGATCATGCGCCCTACGACGGCATGAACCGCGCGCTC	240	
DB	181	CGACTGCTCGTTTACAGCGGGGCCGATCATGCGCCCTACGACGGCATGAACCGCGCGCTC	240	
QY	241	GGGTGAGCCACAGGCGAATGGGTACTTTTTTAGCGCGCGACGACACCTCTACGAACCA	300	
DB	241	GGGTGAGCCACAGGCGAATGGGTACTTTTTTAGCGCGCGACGACACCTCTACGAACCA	300	
QY	301	ACCACTGGCGCCAGGTAGCGCTTTCTCGCGGACCATGCGGCGACGCACTCTGCTCTAT	360	
DB	301	ACCACTGGCGCCAGGTAGCGCTTTCTCGCGGACCATGCGGCGACGCACTCTGCTCTAT	360	
QY	361	GGCGATGTTGTGATGCGTTTCGACGAAAAAGCCGGATGCGCGACCTTCGACCTCGACCGC	420	
DB	361	GGCGATGTTGTGATGCGTTTCGACGAAAAAGCCGGATGCGCGACCTTCGACCTCGACCGC	420	
QY	421	CTCCTATTTGAGACGAAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	480	
DB	421	CTCCTATTTGAGACGAAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	480	
QY	481	GGCATCGGCCCTTACAACTCGGCTACCGAGTCTGGGGGACGTCGGGACCTTCAATATTCG	540	
DB	481	GGCATCGGCCCTTACAACTCGGCTACCGAGTCTGGGGGACGTCGGGACCTTCAATATTCG	540	
QY	541	TGCTTCTCCAAACCGCGCGTGATTAACCGCTACATGACGTCGTGATTTCCGAATATCAAC	600	
DB	541	TGCTTCTCCAAACCGCGCGTGATTAACCGCTACATGACGTCGTGATTTCCGAATATCAAC	600	
QY	601	GACATGACCGGTTTACGATGAGGACGGGACTGATAAAGAGTTTCAGAAAACGGCTGCCA	660	
DB	601	GACATGACCGGTTTACGATGAGGACGGGACTGATAAAGAGTTTCAGAAAACGGCTGCCA	660	
QY	661	ATGTACTTCTGGGTTGCAAGGTGGGAGACTTTCGAGCGGCATGCTGGCGTTTTTGAAGAC	720	
DB	661	ATGTACTTCTGGGTTGCAAGGTGGGAGACTTTCGAGCGGCATGCTGGCGTTTTTGAAGAC	720	
QY	721	AAGGGAATTCGCGCTCTGCGCTTTCGCTACGCGGTTGATGAGGGTTTAAGCGCGCTCTCAAA	780	

|||||
721 AAGGAGAAATCCCGCTCTGCGCTTGGCTACGCGGTGATAGGGTTAAGCGCTTCCAAA 780
781 GAACGAAGCGCAGAACCGTAG 801
781 GAACGAAGCGCAGAACCGTAG 801
RESULT 2
AAT74479
ID AAT74479 standard; DNA; 4435 BP.
XX AC AAT74479;
XX
XX 17-OCT-2003 (revised)
XX 25-MAR-2003 (revised)
XX 11-MAR-1998 (first entry)
XX
XX Part of the GS region DNA sequence from M. paratuberculosis.
XX
XX GS; pathogenesis island; pathogenic protein; mycobacterial disease;
XX cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
XX vaccine; inflammatory disease; sarcoidosis; Johne's disease; ss.
XX
XX Mycobacterium avium subsp. paratuberculosis.
XX
XX Key Location/Qualifiers
XX CDS 201..1232
XX FT /*tag= a
XX FT /note= "encodes AAW21769"
XX FT CDS 1172..2191
XX FT /*tag= b
XX FT /note= "encodes AAW21771"
XX FT 2467..3189
XX FT /*tag= c
XX FT /note= "encodes AAW21773"
XX FT CDS 3335..4135
XX FT /*tag= d
XX FT /note= "encodes AAW21775"
XX
XX WO9723624-A2.
XX
XX 03-JUL-1997.
XX
XX 23-DEC-1996; 96WO-GB003221.
XX
XX 21-DEC-1995; 95GB-00026178.
XX
XX (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.
XX
XX Hermontaylor J, Doran T, Millar D, Tizard M, Loughlin M, Sumar N;
XX Ford J;
XX
XX WPI; 1997-351061/32.
XX
XX New isolated pathogenicity island from mycobacteria - used to develop
XX products for detection, diagnosis, prevention and treatment of
XX mycobacteria infections.
XX
XX Claim 5; Page 43-44; 62pp; English.
XX
XX The present sequence represents a novel polynucleotide sequence
XX designated "GS". GS is a pathogenicity island of 8 kb of DNA comprising a
XX core region of 5.75 kb with multiple open reading frames (ORFs) and an
XX adjacent transmissible element of 2.5 kb. The ORFs, and also the
XX transmissible element, encode proteins which may be linked to
XX pathogenesis, such as providing receptors for cellular recognition. GS
XX was discovered and characterised using differential DNA analysis
XX technology. It is found within Mycobacterium paratuberculosis and it has
XX also been identified in Mycobacterium avium subspecies silvaticum. These
XX pathogenic mycobacteria cause chronic inflammation of the intestine and
XX Crohn's disease in humans. The protein products of the ORFs of GS can be
XX used for detecting mycobacteria or for diagnosing, treating or preventing
XX

CC mycobacterial disease. In particular they can be used as vaccines for
CC inflammatory diseases such as Crohn's disease or sarcoidosis in humans or
CC Johne's disease in animals. (Updated on 25-MAR-2003 to correct PI field.)
CC (Updated on 17-OCT-2003 to standardise OS field)

SQ Sequence 4435 BP; 904 A; 1289 C; 1288 G; 954 T; 0 U; 0 Other;

Query Match 100.0%; Score 801; DB 2; Length 4435;
Best Local Similarity 100.0%; Pred. No. 1.2e-221;
Matches 801; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACTGGCCAGTGTTCGATATTAATCCCTACCTTCATGCGAGCGTGACGTGCAA 60
DB 3335 ATGACTGGCCAGTGTTCGATATTAATCCCTACCTTCATGCGAGCGTGACGTGCAA 3394
QY 61 GCCTGCCTCGGAAGCATCGTCGGGAGACCTACCGGGAAGTGAAGTGTCTTCTGCGAC 120
DB 3395 GCCTGCCTCGGAAGCATCGTCGGGAGACCTACCGGGAAGTGAAGTGTCTTCTGCGAC 3454
QY 121 GCGGTTTCGACCGATCGGACCTCGACATCGGAACAGTTCCTCGCCCGGAACCTCGGCTCG 180
DB 3455 GCGGTTTCGACCGATCGGACCTCGACATCGGAACAGTTCCTCGCCCGGAACCTCGGCTCG 3514
QY 181 CGACTGCTTCACAGCGGCGCGATGATGCGCCCTACGAGCCATGAACCGCGGCTC 240
DB 3515 CGACTGCTTCACAGCGGCGCGATGATGCGCCCTACGAGCCATGAACCGCGGCTC 3574
QY 241 GCGGTAGCCACAGGCGAATGGGTACTTTTATAGCGCGCGACACACCTCTACGAACA 300
DB 3575 GCGGTAGCCACAGGCGAATGGGTACTTTTATAGCGCGCGACACACCTCTACGAACA 3634
QY 301 ACCAGTTGGCCAGTAGACCGCTTTCTCGCGACCATGCGGCAAGCATCTTCTCTAT 360
DB 3635 ACCAGTTGGCCAGTAGACCGCTTTCTCGCGACCATGCGGCAAGCATCTTCTCTAT 3694
QY 361 GCGATGTTGATCGGTTGCGAAGAAAGCGGCGATGCGGACCTTTGCACTCGACCGC 420
DB 3695 GCGATGTTGATCGGTTGCGAAGAAAGCGGCGATGCGGACCTTTGCACTCGACCGC 3754
QY 421 CTCCTATTGAGACGAATTTGCGCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
DB 3755 CTCCTATTGAGACGAATTTGCGCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 3814
QY 481 GGCATCGGCGCTTACAACTCGGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTCG 540
DB 3815 GGCATCGGCGCTTACAACTCGGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTCG 3874
QY 541 TGCTTCTCAACCGCGCTGATTAACCGCTACATGAGCGTGTGATTTCCGATACAAAC 600
DB 3875 TGCTTCTCAACCGCGCTGATTAACCGCTACATGAGCGTGTGATTTCCGATACAAAC 3934
QY 601 GACATGACCGGCTTCAGCATGAGGCGAGGACTGATAAGAGTTTCAGAAAACGGCTGCA 660
DB 3935 GACATGACCGGCTTCAGCATGAGGCGAGGACTGATAAGAGTTTCAGAAAACGGCTGCA 3994
QY 661 ATGTACTTCTGCGTTGCGAGGTGGGAGACTTCGAGCGCATGCTGCGGTTTTTGAAGAC 720
DB 3995 ATGTACTTCTGCGTTGCGAGGTGGGAGACTTCGAGCGCATGCTGCGGTTTTTGAAGAC 4054
QY 721 AAGGAGATTCGCGCTTCGCGCTTCGATGAGCGGTTGATAAGGTTTAAAGCGCTTCGAAA 780
DB 4055 AAGGAGATTCGCGCTTCGCGCTTCGATGAGCGGTTGATAAGGTTTAAAGCGCTTCGAAA 4114
QY 781 GAACGAAGCGCAGAACCGTAG 801
DB 4115 GAACGAAGCGCAGAACCGTAG 4135

RESULT 3

AAT74469

ID AAT74469 standard; DNA; 801 BP.

XX

AC AAT74469;

XX

DT 17-OCT-2003 (revised)

DT 25-MAR-2003 (revised)

DT 11-MAR-1998 (first entry)

XX

DE Open reading frame F from GS region of *M. avium* subspecies *silvaticum*.

XX

KW GS; pathogenicity island; pathogenic protein; mycobacterial disease;
KW cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
KW vaccine; inflammatory disease; sarcoidosis; Johne's disease; ss.

XX

OS *Mycobacterium avium*; subspecies *silvaticum*.

XX

PH Key Location/Qualifiers

FT CDS 1..801

FT /*tag= a

XX

PN W09723624-A2.

XX

PD 03-JUL-1997.

XX

PF 23-DEC-1996; 96WO-GB003221.

XX

PR 21-DEC-1995; 95GB-00026178.

XX

PA (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.

XX

PI Hermontaylor J, Doran T, Millar D, Tizard M, Loughlin M, Sumar N;

PI Ford J;

XX

DR WPI; 1997-351061/32.

XX

P-PSDB; AAW21774.

DR

PT New isolated pathogenicity island from mycobacteria - used to develop
PT products for detection, diagnosis, prevention and treatment of
PT mycobacteria infections.

XX

PS Claim 6; Page 50; 62pp; English.

XX

CC The present sequence represents one of eight open reading frames (ORFs)
CC of a novel polynucleotide sequence designated "GS". GS is a pathogenicity
CC island of 8 kb of DNA comprising a core region of 5.75 kb with multiple
CC ORFs and an adjacent transmissible element of 2.5 kb. The ORFs, and also
CC the transmissible element, encode proteins which may be linked to
CC pathogenicity, such as providing receptors for cellular recognition. GS
CC was discovered and characterised using differential DNA analysis
CC technology. It is found within *Mycobacterium paratuberculosis* and it has
CC also been identified in *Mycobacterium avium* subspecies *silvaticum*. These
CC pathogenic mycobacteria cause chronic inflammation of the intestine and
CC Crohn's disease in humans. The protein products of the ORFs of GS can be
CC used for detecting mycobacteria or for diagnosing, treating or preventing
CC mycobacterial disease. In particular they can be used as vaccines for
CC inflammatory diseases such as Crohn's disease or sarcoidosis in humans or
CC Johne's disease in animals. (Updated on 25-MAR-2003 to correct PI field.)
CC (Updated on 17-OCT-2003 to standardise OS field)

XX

SQ Sequence 801 BP; 167 A; 232 C; 229 G; 173 T; 0 U; 0 Other;

Query Match

Best Local Similarity 99.8%; Score 799.4; DB 2; Length 801;

Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGACTGGCCAGTGTTCGATATTAATCCCTACCTTCATGCGAGCGTGACGTGCAA 60

DB 1 ATGACTGGCCAGTGTTCGATATTAATCCCTACCTTCATGCGAGCGTGACGTGCAA 60

QY 61 GCCTGCCTCGGAAGCATCGTCGGGAGACCTACCGGGAAGTGAAGTGTCTTCTGCGAC 120

DB 61 GCCTGCCTCGGAAGCATCGTCGGGAGACCTACCGGGAAGTGAAGTGTCTTCTGCGAC 120

QY 121 GCGGTTTCGACCGATCGGACCTCGACATCGGAACAGTTCGCGCCCGGAACCTCGGCTCG 180

DB 121 GCGGTTTCGACCGATCGGACCTCGACATCGGAACAGTTCGCGCCCGGAACCTCGGCTCG 180

```
QY 181 CGACTGGTCTTCAAGCGGGCCCGATGATGGCCCTTACGAGCGCCATGAACGGCGGCTC 240
Db 181 CGACTGGTCTTCAAGCGGGCCCGATGATGGCCCTTACGAGCGCCATGAACGGCGGCTC 240
QY 241 GCGTAGCCACAGGCGAATGGGTACTTTTATAGCGCGCGAGCACACCTCTACGAACCA 300
Db 241 GCGTAGCCACAGGCGAATGGGTACTTTTATAGCGCGCGAGCACACCTCTACGAACCA 300
QY 301 ACCAGTTGGCCAGGTAGCGCTTTTCTCGCGACCATGCGGCAAGCCATCTTGTCTAT 360
Db 301 ACCAGTTGGCCAGGTAGCGCTTTTCTCGCGACCATGCGGCAAGCCATCTTGTCTAT 360
QY 361 GCGCATGTTGTATCGTTCGACGAAAGCCGCGATGCGGACCTTTCGACCTCGACGCG 420
Db 361 GCGCATGTTGTATCGTTCGACGAAAGCCGCGATGCGGACCTTTCGACCTCGACGCG 420
QY 421 CTCCTATTGAGACGAAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
Db 421 CTCCTATTGAGACGAAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
QY 481 GGCATCGGCCCTTACAACCTCGCTACCGAGTCTGGCGGACTGGGACTTCAATATTGCG 540
Db 481 GGCATCGGCCCTTACAACCTCGCTACCGAGTCTGGCGGACTGGGACTTCAATATTGCG 540
QY 541 TGCCTCTCGAACCCGCGCTGATTAACCGCTACATGAGCGTGTGATTTCCGAATACAA 600
Db 541 TGCCTCTCGAACCCGCGCTGATTAACCGCTACATGAGCGTGTGATTTCCGAATACAA 600
QY 601 GACATGACCGGCTTCAGATGAGGAGGAGGACTGATAAGAGTTTCAGAAAACGGCTGCCA 660
Db 601 GACATGACCGGCTTCAGATGAGGAGGAGGACTGATAAGAGTTTCAGAAAACGGCTGCCA 660
QY 661 ATGTACTTCTGGGTTGCGAGGTTGGAGACTTTCAGCGCATGCTGGCGTTTGAAGAC 720
Db 661 ATGTACTTCTGGGTTGCGAGGTTGGAGACTTTCAGCGCATGCTGGCGTTTGAAGAC 720
QY 721 AAGGAGATCGCGCTGCGCTTGGCTAGCGGTTGATAGGGTTAAGCGCGTCTCCAAA 780
Db 721 AAGGAGATCGCGCTGCGCTTGGCTAGCGGTTGATAGGGTTAAGCGCGTCTCCAAA 780
QY 781 GAACGAAGCGCAACCGTAG 801
Db 781 GAACGAAGCGCAACCGTAG 801

RESULT 4
ID AAT74478
XX AAT74478 standard; DNA; 7995 BP.
AC AAT74478;
XX
DT 17-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 11-MAR-1998 (first entry)
XX
DE Complete GS region DNA sequence from M. avium subspecies silvaticum.
XX
KW GS; pathogenecity island; pathogenic protein; mycobacterial disease;
KW cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
KW vaccine; inflammatory disease; sarcoidosis; Johne's disease; ss.
XX
OS Mycobacterium avium; subspecies silvaticum.
XX
FH Key
FT Location/Qualifiers
FT 50..427
FT /tag= a
FT /note= "encodes AAW21766"
FT 772..1605
FT /tag= b
FT /note= "encodes AAW21767"
FT 1814..2845
FT /tag= c
```

```
FT CDS /note= "encodes AAW21768"
FT 2785..3804
FT /*tag= d
FT /note= "encodes AAW21770"
FT 4080..4802
FT /*tag= e
FT /note= "encodes AAW21772"
FT 4947..5747
FT /*tag= f
FT /note= "encodes AAW21774"
FT 6176..7042
FT /*tag= g
FT /note= "encodes AAW21776"
FT complement(6215..7953)
FT /*tag= h
FT /note= "encodes AAW21777-78"
XX
XX WO9723624-A2.
XX
XX 03-JUL-1997.
XX
XX 23-DEC-1996; 96WO-GB003221.
XX
XX 21-DEC-1995; 95GB-00026178.
XX
XX (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.
XX
XX Hermon Taylor J, Doran T, Millar D, Tizard M, Loughlin M, Sumar N;
XX Ford J;
XX
XX WPI, 1997-351061/32.
XX
XX New isolated pathogenecity island from mycobacteria - used to develop
XX products for detection, diagnosis, prevention and treatment of
XX mycobacteria infections.
XX
XX Claim 5; Page 40-43; 62pp; English.
XX
XX The present sequence represents a novel polynucleotide sequence
XX designated "GS". GS is a pathogenecity island of 8 kb of DNA comprising a
XX core region of 5.75 kb with multiple open reading frames (ORFs) and an
XX adjacent transmissible element of 2.5 kb. The ORFs, and also the
XX transmissible element, encode proteins which may be linked to
XX pathogenecity, such as providing receptors for cellular recognition. GS
XX was discovered and characterised using differential DNA analysis
XX technology. It is found within Mycobacterium avium subspecies silvaticum
XX and it has also been identified in Mycobacterium paratuberculosis. These
XX pathogenic mycobacteria cause chronic inflammation of the intestine and
XX Crohn's disease in humans. The protein products of the ORFs of GS can be
XX used for detecting mycobacteria or for diagnosing, treating or preventing
XX mycobacterial disease. In particular they can be used as vaccines for
XX inflammatory diseases such as Crohn's disease or sarcoidosis in humans or
XX Johne's disease in animals. (Updated on 25-MAR-2003 to correct PI field.)
XX (Updated on 17-OCT-2003 to standardise OS field)
XX
XX Sequence 7995 BP; 1574 A; 2323 C; 2462 G; 1636 T; 0 U; 0 Other;
XX
XX Query Match 99.8%; Score 799.4; DB 2; Length 7995;
XX Best Local Similarity 99.9%; Pred. No. 4.2e-221;
XX Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGACTGCGCCAGTGTCTCGATTAATTCCTACCTTCAATGCGAGCGGTGACGCTGCAA 60
Db 4947 ATGACTGCGCCAGTGTCTCGATTAATTCCTACCTTCAATGCGAGCGGTGACGCTGCAA 5006
QY 61 GCCTGCCTCGGAAGCATGCTCGCGGAGACCTACCGGGAAGTGGAGTCTCTTCGAC 120
Db 5007 GCCTGCCTCGGAAGCATGCTCGCGGAGACCTACCGGGAAGTGGAGTCTCTTCGAC 5066
QY 121 GCGGTTTCGACCGGATCGGACCTCGACATCGGAAACAGTTTCGCGCCCGGAACCTCGGCTCG 180
Db 5067 GCGGTTTCGACCGGATCGGACCTCGACATCGGAAACAGTTTCGCGCCCGGAACCTCGGCTCG 5126
```


ACA40754; ' 1
 19-JUN-2003 (first entry)
 Prokaryotic essential gene #22411.
 Antisense; ds; prokaryotic essential gene; cell proliferation;
 drug design; gene.
 Mycobacterium tuberculosis.
 WO200277183-A2.
 03-OCT-2002.
 21-MAR-2002; 2002WO-US009107.
 21-MAR-2001; 2001US-00815242.
 06-SEP-2001; 2001US-00948993.
 25-OCT-2001; 2001US-0342923P.
 08-FEB-2002; 2002US-00072851.
 06-MAR-2002; 2002US-0362699P.
 (ELIT-) ELITRA PHARM INC.
 Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 WPI; 2003-029926/02.
 P-PSDB; ABU36884.
 New antisense nucleic acids, useful for identifying proteins or screening
 for homologous nucleic acids required for cellular proliferation to
 isolate candidate molecules for rational drug discovery programs.
 Claim 14; SEQ ID NO 28624; 1766pp; English.
 The invention relates to an isolated nucleic acid comprising any one of
 the 6213 antisense sequences given in the specification where expression
 of the nucleic acid inhibits proliferation of a cell. Also included are:
 (1) a vector comprising a promoter operably linked to the nucleic acid
 encoding a polypeptide whose expression is inhibited by the antisense
 nucleic acid; (2) a host cell containing the vector; (3) an isolated
 polypeptide or its fragment whose expression is inhibited by the
 antisense nucleic acid; (4) an antibody capable of specifically binding
 the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 proliferation or the activity of a gene in an operon required for
 proliferation; (7) identifying a compound that influences the activity of
 the gene product or that has an activity against a biological pathway
 required for proliferation, or that inhibits cellular proliferation; (8)
 identifying a gene required for cellular proliferation or the biological
 pathway in which a proliferation-required gene or its gene product lies
 or a gene on which the test compound that inhibits proliferation of an
 organism acts; (9) manufacturing an antibiotic; (10) profiling a
 compound's activity; (11) a culture comprising strains in which the gene
 product is overexpressed or underexpressed; (12) determining the extent
 to which each of the strains is present in a culture or collection of
 strains; or (13) identifying the target of a compound that inhibits the
 proliferation of an organism. The antisense nucleic acids are useful for
 identifying proteins or screening for homologous nucleic acids required
 for cellular proliferation to isolate candidate molecules for rational
 drug discovery programs, or for screening homologous nucleic acids
 required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
K. pneumoniae or *P. aeruginosa*. The present sequence is one of the target
 prokaryotic essential genes. Note: The sequence data for this patent did
 not form part of the printed specification, but was obtained in
 electronic format directly from WIPO at
 ftp.wipo.int/pub/published_pct_sequences

Query Match 43.2%; Score 346; DB 8; Length 828;
 Best Local Similarity 71.0%; Pred. NO. 7e-90;

Sequence 828 BP; 162 A; 264 C; 233 G; 169 T; 0 U; 0 Other;

Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;
 QY 1 ATGACTGGCCAGTGTCTTCGATAATATCTCTTCAATGACAGCGGTGACGCTGCAA 60
 DB 58 ATGGCGGACCAATGTTTTTCGATCATCATCCACCTTGAACGTGGCTGCGGTATGCGT 117
 QY 61 GCCTGCTCGGAAGCATCGTCGGGCAGACCTACCCGGGAAGTGGAGTGGTCTTCTGCGAC 120
 DB 118 GCCTGCTCGGACAGCATCGCCCGTCAGACCTGCGGTGACTTCGAGCTGGTACTGCTCGAC 177
 QY 121 GCGGTTTCGACCGATCGGACCTCGACATCGGAACAGTTCGCGCCCGGAACCTCGGCTCG 180
 DB 178 GCGGCTCGACGSAAGAAACCTTCGACATCGCAATTTTCGCCCCCAACCTTCGCGCGAG 237
 QY 181 CGACTGGTCTGTTACAGCGGCGCCGATGATGGCCCTACGACGCGCATGAACCGCGGCTC 240
 DB 238 CGTTGATCATTCATCGGACACCGACCGAGGCGTCTACGACGCGCATGAACCGGCGGTG 297
 QY 241 GCGGTAGCCACAGCGGAATGGGTACTTTTTTATAGCGCGCGACGACACCTCTAGAACCA 300
 DB 298 GACCTGGCCACCGGAAACGTGGTGTCTTTCTGGGCGCGGACGACAGCTGTACGAGGCT 357
 QY 301 ACCAGTTGGCCAGTAGCCGCTTTCTCGGCGACCATGCGGCGACCATCTGTCTAT 360
 DB 358 GACACCTGGCGGCGGTGGCGCTTCATTGGCGAAACACGAGCCGCGATCTGGTATAT 417
 QY 361 GCGGATGTTGATGCGTTCGACGAAAGCCGCGATGCGCGACCTTCGACCTCGACCGC 420
 DB 418 GCGGCGTGTATGCGCTCAACCAATTCGCTGGGCGGCTTCGACCTCGACCGT 477
 QY 421 CTCCTATTGAGACGAATTTGTGCCCAATTCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
 DB 478 CTGTTGTTCAAGCGCAACATCTGCCATCAGCGCATCTTCTACCGCGGAGCTCTTCGCG 537
 QY 481 GGCATCGGCTTACAACTCGGCTACCGATCGGAGTCTGGGCGGAGCTGGAGCTTCAATTCGC 540
 DB 538 ACCATCGGCTTCAAACTCGGCTACCGGCTCGGCGGCTTCGCGGCTTCGAGCTTCGCG 597
 QY 541 TGCCTTCCAAACCGCGGCTGATTTACCGCTACATGACGCTCGTGTATTTCCGAATACAA 600
 DB 598 TGCCTTCCAAACCGCGGCTGCTCACCCTGCTACATGACGCTGGTGTTCGAAGCTACAA 657
 QY 601 GACATGACCGGCTTCAGCATGAGCGGAGCTGATAAAGAGTTTCAGAAAAACGGTGCCT 660
 DB 658 GAATTCGCGGCGCTCAGCAATACGATCG---TCGACAAGGAGTTTTCGAGCGGCTGCGG 714
 QY 661 ATGTAC 666
 DB 715 ATGTCC 720

RESULT 8
 ADB74252/C
 ID ADB74252 standard; DNA; 32155 BP.
 XX AC ADB74252;
 XX DT 04-DEC-2003 (first entry)
 XX DB Mycobacterium tuberculosis DNA.
 XX KW Non-naturally occurring peptide; anion pump protein; tuberculosis;
 XX KM hypersensitivity reaction; tuberculostatic; gene; ds.
 XX OS Mycobacterium tuberculosis.
 XX PN US6583266-B1.
 XX PD 24-JUN-2003.
 XX PF 16-SEP-1994; 94US-00311731.
 XX PR 19-AUG-1993; 93US-00109181.

PR 22-OCT-1993; 93US-00142558.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Smith DR, Mao J;
XX WPI; 2003-656441/62.
DR P-PSDB; ADB74253, ADB74254, ADB74255, ADB74256, ADB74257, ADB74258,
DR ADB74259, ADB74260, ADB74261, ADB74262, ADB74263, ADB74264, ADB74265,
DR ADB74266, ADB74267, ADB74268, ADB74269, ADB74270, ADB74271, ADB74272,
DR ADB74273, ADB74369, ADB74370.
XX New Mycobacterium tuberculosis anion pump peptide useful for as
PT tuberculosis vaccine and diagnosis of tuberculosis infection.
XX
XX Disclosure; SEQ ID NO 1; 26pp; English.
XX The invention relates to a non-naturally occurring peptide of
CC Mycobacterium tuberculosis comprising an amino acid sequence
CC corresponding to an anion pump protein. The invention also relates to a
CC non-naturally occurring nucleic acid corresponding to a DNA sequence of
CC Mycobacterium tuberculosis or Mycobacterium leprae. The new peptide is
CC useful as a vaccine against Mycobacterium tuberculosis or Mycobacterium
CC leprae or for screening for new tuberculosis drugs. Purified proteins
CC derived from the sequences of the invention may elicit a specific immune
CC response. The peptide may also be used to detect hypersensitivity
CC reactions of individuals exposed to Mycobacterium tuberculosis or
CC Mycobacterium leprae. The proteins and peptides may be affixed to solid
CC supports to detect antibodies typical of hypersensitivity reactions, from
CC a patient's sera. This sequence represents Mycobacterium tuberculosis DNA
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification but was obtained in electronic format
CC directly from USPIO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 32155 BP; 5716 A; 9444 C; 10766 G; 6229 T; 0 U; 0 Other;
Query Match 43.2%; Score 346; DB 10; Length 32155;
Best Local Similarity 71.0%; Pred. No. 2.7e-89;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;
QY 1 ATGACTGGCCAGTGTTCGATGATATTCCTTACCTTCAATGACGGGTGACGCTGCAA 60
Db 15076 ATGGCGCACCAATGTTTCGATCATATCCACCTTGAAGCTGGCTGCGTATTGCT 15017
QY 61 GCCTGCTCGAAGCATGTCGGGAGACCTACCGGGAAGTGAAGTGGTCTTGTGAC 120
Db 15016 GCCTGCTCGACAGCATCGCCGTGAGACCTGCGGTGACTTCGAGCTGGTACTGTCGAC 14957
QY 121 GCGGTTTCGACCGATCGGACCTCGACATCGGACAGTTCCGCCCGGAACTCGGCTCG 180
Db 14956 GCGGCTCGACGAGCAACCTTCGACATCGCAATTCGCCCCCAACCTTCGGCGAG 14897
QY 181 CGACTGTGCTTTCACAGCGGGCCCGATGATGGCCCTACGACGGCATGAACCGGGCTC 240
Db 14896 CGGTTGATCATTCATCGGACACCGACCGAGCGGTCTAGACGCCATGAACCGGGGTG 14837
QY 241 GCGGTAGCCACAGGCAATGGGTACTTTTATGCGCGCGAGCAGACCTCTACGAACCA 300
Db 14836 GACCTGGCCACCGGAACTGGTGTCTTTCTGCGCGCGGAGCAGCAGCTGTGAGGCT 14777
QY 301 ACCAGTGGCCAGTAGCGCTTTTCTCGCGACCATGCGGAAGCATCTGTCTAT 360
Db 14776 GACACCTTGGCGGGTGGCCCTTCAITGGGAAACAGAGCCAGCATCTGGTATAT 14717
QY 361 GCGCATGTTGTGATCGGTTCGACGAAAGCCGGCATGCGGACCTTTTCGACCTCGACCGC 420
Db 14716 GCGCAGGTGATCATCGCTCAACCAATTCGCTGGGTGGCGCTTCGACCTCGACCGT 14657
QY 421 CTCCTATTGTGACAGCAATTTGGCCACCAATCGATCTTTACCGCGGTGAGCTTTTCAC 480
Db 14656 CTGTTGTTTCAAGCGCAACATCTGCCATCAGGCGATCTTTCTACCGCGGAGCTCTTCGGC 14597
QY 481 GGCATCGGCCCTTACAACTCGGCTACCGAGTCTGGCGGAGTGGGACTTCAATATTGCG 540

Db 14596 ACCATCGTCCCTACAACTCCGCTACCGGCTCTGCGGACTGGGACTTCAATATTGCG 14537
QY 541 TGCCTTCCAAACCCGGCGCTGATTACCGGTACATGAGCGTGTGATTTCCGAATACAAAC 600
Db 14536 TGCCTTCCAAACCCGGCGCTGTCACCGCTTACATGACAGCTGCTGTTGCAAGCTACAAAC 14477
QY 601 GACATGACCGGCTTCACGNTGAGCGAGGAGTCTATAAAGAGTTCAGAAAACGGCTGCCA 660
Db 14476 GAAATCGCGGGCTCAGCAATACGATCG---TCGACAAAGGAGTTTTGAAGCGGCTGCCG 14420
QY 661 ATGTAC 666
Db 14419 ATGTCC 14414
RESULT 9
AAI99682_33
Continuation (34 of 45) of AAI99682 from base 3300001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682
WP Fragment Name Begin End
WP AAI99682_00 1 110000
WP AAI99682_01 100001 210000
WP AAI99682_02 200001 310000
WP AAI99682_03 300001 410000
WP AAI99682_04 400001 510000
WP AAI99682_05 500001 610000
WP AAI99682_06 600001 710000
WP AAI99682_07 700001 810000
WP AAI99682_08 800001 910000
WP AAI99682_09 900001 1010000
WP AAI99682_10 1000001 1110000
WP AAI99682_11 1100001 1210000
WP AAI99682_12 1200001 1310000
WP AAI99682_13 1300001 1410000
WP AAI99682_14 1400001 1510000
WP AAI99682_15 1500001 1610000
WP AAI99682_16 1600001 1710000
WP AAI99682_17 1700001 1810000
WP AAI99682_18 1800001 1910000
WP AAI99682_19 1900001 2010000
WP AAI99682_20 2000001 2110000
WP AAI99682_21 2100001 2210000
WP AAI99682_22 2200001 2310000
WP AAI99682_23 2300001 2410000
WP AAI99682_24 2400001 2510000
WP AAI99682_25 2500001 2610000
WP AAI99682_26 2600001 2710000
WP AAI99682_27 2700001 2810000
WP AAI99682_28 2800001 2910000
WP AAI99682_29 2900001 3010000
WP AAI99682_30 3000001 3110000
WP AAI99682_31 3100001 3210000
WP AAI99682_32 3200001 3310000
WP AAI99682_33 3300001 3410000
WP AAI99682_34 3400001 3510000
WP AAI99682_35 3500001 3610000
WP AAI99682_36 3600001 3710000
WP AAI99682_37 3700001 3810000
WP AAI99682_38 3800001 3910000
WP AAI99682_39 3900001 4010000
WP AAI99682_40 4000001 4110000
WP AAI99682_41 4100001 4210000
WP AAI99682_42 4200001 4310000
WP AAI99682_43 4300001 4410000
WP AAI99682_44 4400001 4411529
Query Match 43.2%; Score 346; DB 4; Length 110000;
Best Local Similarity 71.0%; Pred. No. 4.2e-89;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;
QY 1 ATGACTGGCCAGTGTTCGATGATATTCCTTACCTTCAATGACGGGTGACGCTGCAA 60

Db	104455	ATGTCC	104460		
RESULT 11					
AAI199683_33					
Continuation (34 of 44) of AAI199683 from base 3300001 (Mycobacterium tuberculosis strain					
WP Sequence split into 44 fragments LOCUS AAI199683 Accession Aai199683					
WP	Fragment Name	Begin	End		
WP	AAI199683_00	1	110000		
WP	AAI199683_01	100001	210000		
WP	AAI199683_02	200001	310000		
WP	AAI199683_03	300001	410000		
WP	AAI199683_04	400001	510000		
WP	AAI199683_05	500001	610000		
WP	AAI199683_06	600001	710000		
WP	AAI199683_07	700001	810000		
WP	AAI199683_08	800001	910000		
WP	AAI199683_09	900001	1010000		
WP	AAI199683_10	1000001	1110000		
WP	AAI199683_11	1100001	1210000		
WP	AAI199683_12	1200001	1310000		
WP	AAI199683_13	1300001	1410000		
WP	AAI199683_14	1400001	1510000		
WP	AAI199683_15	1500001	1610000		
WP	AAI199683_16	1600001	1710000		
WP	AAI199683_17	1700001	1810000		
WP	AAI199683_18	1800001	1910000		
WP	AAI199683_19	1900001	2010000		
WP	AAI199683_20	2000001	2110000		
WP	AAI199683_21	2100001	2210000		
WP	AAI199683_22	2200001	2310000		
WP	AAI199683_23	2300001	2410000		
WP	AAI199683_24	2400001	2510000		
WP	AAI199683_25	2500001	2610000		
WP	AAI199683_26	2600001	2710000		
WP	AAI199683_27	2700001	2810000		
WP	AAI199683_28	2800001	2910000		
WP	AAI199683_29	2900001	3010000		
WP	AAI199683_30	3000001	3110000		
WP	AAI199683_31	3100001	3210000		
WP	AAI199683_32	3200001	3310000		
WP	AAI199683_33	3300001	3410000		
WP	AAI199683_34	3400001	3510000		
WP	AAI199683_35	3500001	3610000		
WP	AAI199683_36	3600001	3710000		
WP	AAI199683_37	3700001	3810000		
WP	AAI199683_38	3800001	3910000		
WP	AAI199683_39	3900001	4010000		
WP	AAI199683_40	4000001	4110000		
WP	AAI199683_41	4100001	4210000		
WP	AAI199683_42	4200001	4310000		
WP	AAI199683_43	4300001	4403765		
Query Match 43.2%; Score 346; DB 4; Length 110000;					
Best Local Similarity 71.0%; Pred. No. 4,2e-89;					
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;					
Qy	1	ATGACTGCGCAGTGTTCGATATATTCCTACCTTCAATGACGCGGTGACGCTGCAA	60		
Db	3798	ATGGCGGCACCAATGTTTTTCGATCATATCCACCTTGAACGTGCGGTATTGCCT	3857		
Qy	61	GCCTGCTCGAAGCATGTCGGGAGACCTACCGGGAAGTGGAGTGTCTTGTCTGAC	120		
Db	3858	GCCTGCTCGACAGCATGCCCGGTGACCTTGCAGCTGGTACTGGTTCGAC	3917		
Qy	121	GCGGTTTCACCGATCGGACCTCGACATCGGAACAGTTTTCCGCCCGGAACCTCGGCTCG	180		
Db	3918	GCGGCTCGACGGACGAACCTTCGACATCGCCACATTTTCGCCCCCAACCTCGGCGAG	3977		
Qy	181	CGACTGGTTCGTTACAGCGGGGCCGATGATGGCCCCCTACGACGCGCATGAACCGCGGCTC	240		
Db	3978	CGGTTGATCATTCATCGCGACACCGACCGGCGCTTACGACGCGCATGAACCGCGCGGTG	4037		
Qy	241	GGCGTAGCCACAGCGAATGGGTACTTTTTTAGGCGCGCAGACACCCCTCTACGAACCA	300		
Db	4038	GACCTGGCCACCGGAACGTGTGCTCTTTCTGGCGCGGACGACAGCTGTACGAGCT	4097		
Qy	301	ACCAAGTTGGCCAGGTAGCGCTTTTTCTCGGCGACCATGCGGCAAGCCATCTTGTCTAT	360		
Db	4098	GACACCCCTGGCGCGGTGGCGCCTTCATTGGCGAAACACGAGCCACGATCTGGTATAT	4157		
Qy	361	GGCGATGTTGTGATGCGTTTCACGAAAAAGCGGATGCGGACCTTTTCGACCTCGACCGC	420		
Db	4158	GGCGACGTGATCATGCGCTCAACCAATTCGCTGGGGTGGCGCTTCGACCTCGACCGT	4217		
Qy	421	CTCCTATTGAGACGAATTTGTGCCCAATTCGATCTTTTACCGCGGTGAGCTTTTCGAC	480		
Db	4218	CTGTTGTTCAAGCGCAACATCTGCCATCAGCGGATCTTTACCGCGCGGACTCTTTCGGC	4277		
Qy	481	GGCATCGGCCCTTACAACCTGCGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTGCG	540		
Db	4278	ACCATCGTCCCTACAACCTCCGCTACCGGTCTCTGGCGGACTGGGACTTCAATATTGCG	4337		
Qy	541	TGCTTCTCCAACCGCGGCTGATTTACCGCTACATGAGACGTGCGTATTCGGAATACAAC	600		
Db	4338	TGCTTTTCCAACCCAGCGCTCGTCAACCGCTACATGCACTGCGTGGTTCGAAGCTACA	4397		
Qy	601	GACATGACCGGCTTTCAGCATGAGGACGCGGACTGATAAAGAGTTTCAGAAAAACGGCTGCCA	660		
Db	4398	GAATTCGGCGGCTCAGCAATACGATCG---TCGACAAAGAGTTTTTGAAGCGGCTGCCG	4454		
Qy	661	ATGTAC	666		
Db	4455	ATGTCC	4460		
RESULT 12					
AAI199682_32					
Continuation (33 of 45) of AAI199682 from base 3200001 (Mycobacterium tuberculosis strain					
WP Sequence split into 45 fragments LOCUS AAI199682 Accession Aai199682					
WP	Fragment Name	Begin	End		
WP	AAI199682_00	1	110000		
WP	AAI199682_01	100001	210000		
WP	AAI199682_02	200001	310000		
WP	AAI199682_03	300001	410000		
WP	AAI199682_04	400001	510000		
WP	AAI199682_05	500001	610000		
WP	AAI199682_06	600001	710000		
WP	AAI199682_07	700001	810000		
WP	AAI199682_08	800001	910000		
WP	AAI199682_09	900001	1010000		
WP	AAI199682_10	1000001	1110000		
WP	AAI199682_11	1100001	1210000		
WP	AAI199682_12	1200001	1310000		
WP	AAI199682_13	1300001	1410000		
WP	AAI199682_14	1400001	1510000		
WP	AAI199682_15	1500001	1610000		
WP	AAI199682_16	1600001	1710000		
WP	AAI199682_17	1700001	1810000		
WP	AAI199682_18	1800001	1910000		
WP	AAI199682_19	1900001	2010000		
WP	AAI199682_20	2000001	2110000		
WP	AAI199682_21	2100001	2210000		
WP	AAI199682_22	2200001	2310000		
WP	AAI199682_23	2300001	2410000		
WP	AAI199682_24	2400001	2510000		
WP	AAI199682_25	2500001	2610000		
WP	AAI199682_26	2600001	2710000		
WP	AAI199682_27	2700001	2810000		
WP	AAI199682_28	2800001	2910000		
WP	AAI199682_29	2900001	3010000		
WP	AAI199682_30	3000001	3110000		
WP	AAI199682_31	3100001	3210000		
WP	AAI199682_32	3200001	3310000		
WP	AAI199682_33	3300001	3410000		

WP AAI9682_34 340001 351000
WP AAI9682_35 350001 361000
WP AAI9682_36 360001 371000
WP AAI9682_37 370001 381000
WP AAI9682_38 380001 391000
WP AAI9682_39 390001 401000
WP AAI9682_40 400001 411000
WP AAI9682_41 410001 421000
WP AAI9682_42 420001 431000
WP AAI9682_43 430001 441000
WP AAI9682_44 440001 441159

Query Match 31.2%; Score 249.8; DB 4; Length 110000;
Best Local Similarity 70.2%; Pred. No. 3.5e-61;
Matches 335; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 1 ATGACTGCGCAGTGTCTTCGATATATCCCTACCTTCAATGACAGCGGTGACGTCGAA 60
Db |||||
109524 ATGGCGGACCAATGTTTTCGATCATCTCCACCTTGACGTGGCTGCGGTATGCT 109583
QY 61 GCCTGCTCGGAAGCATCGTGGGCGAGACCTACCGGGAAGTGGAGTGGTCTTGTGAC 120
Db |||||
109584 GCTGCTCGACAGCATCGCCGTGACACCTGCGTGGTCTGAGCTGTGTGTCGAC 109643
QY 121 GCGGTTTCGACGATCGGACCTCGACATCGGGAACAGTTCGCGCCCGGAACCTGCGTGC 180
Db |||||
109644 GCGGCTCGGCGGACCAACCTCGACATCGCAATTTTCGCCCCCAACCTCGCGAG 109703
QY 181 CDACGTGTCGTCACAGGCGGCGCGATGATGCGCCCTACGACGCGCATGACGCGGCTC 240
Db |||||
109704 CGGTGATCATTCATCGGCGACACGACGCGGCGTCTACGACGCGCATGAACCGGCGTG 109763
QY 241 GCGGTAGCCACAGGCGAATGGTACTTTTATGCGCGCGACGACACCTCTACGAACCA 300
Db |||||
109764 GACCTGGCCACGGAACGTGGTGTCTTTCTGGCGGCGGACGACGCTGTACGAGCT 109823
QY 301 ACCAGTTGGCCAGGTAGCGCTTTTCTGCGGACCAATGCGGGAAGCATCTTGTCTAT 360
Db |||||
109824 GACACCTTGGCGGCGGCTTCTATTTGCGGAACACGAGCCCGACGATCTGTATAT 109883
QY 361 GCGGATGTCATGCGTTCGACGAAGACGCGCATGCGGACCTTCGACCTCGACCGC 420
Db |||||
109884 GCGGACGTGATCATCGCTCAACCAATTTCCGCTGGGTGGCGCTTCGACCTCGACCGT 109943
QY 421 CTCCTATTGAGACAAATTTGTGCCCAACATCGATCTTTTACCGCGGTGAGCTTTTC 477
Db |||||
109944 CTGTGTTCAAGCGCAACATCTGCCATCAGGCACTTCTACCGCGCGGACTCTTC 110000

RESULT 13
ACA42153
ID ACA42153 standard; DNA; 900 BP.
XX AC ACA42153;
XX AC ACA42153;
DT 19-JUN-2003 (first entry)
DE Prokaryotic essential gene #23810.
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX OS Pseudomonas aeruginosa.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.

Query Match 7.5%; Score 60; DB 8; Length 900;
Best Local Similarity 50.0%; Pred. No. 7.3e-07;
Matches 209; Conservative 0; Mismatches 200; Indels 9; Gaps 2;

QY 7 GCGCCAGTGTCTCGATAATTTATCCCTACCTTCAATGACGCGGTGACGTCGAAAGCGTGC 66
Db |||||
46 GCGCCGCTGTATCGGTGATCGGCTGCTTCAACGCGGAAAGTACCTCGGAAGAGCC 105
QY 67 CTCGGAAGCATCGTGGGCGAGACCTACCGGAAGTGGAGTGGTCTTGTGACGCGGT 126
Db |||||
106 CTGCGCAGCATCTACGACGAGACTACCCGAATTTTCAAGTGTATCATCTGTCGACGACGT 165
QY 127 TCGACCGATCGGACCTCGACATCGGAACAGTTTCCGCCCGGAACCTCGGCTCGGACTG 186
Db |||||
166 TCCACCGAACACAGCTACGCCATCTCTGAGCAGTTGCAAGAACCTCCAGGTTTCCAGC - 223
QY 187 GTCGTTTCAAGCGGCGCGGATGATGGCCCTTACGACGCGCATGAACCGGCGGTGCGGTA 246
Db |||||
224 ----TCTACCGCAGCAGACACGCGGCTGACGCGGCTGAACTTCGGCTCGGACAC 279
QY 247 GCCACAGCGAATGGGTACTTTTTTTAGGCGCGGACGACACCTCTACGAACCAACACG 306
Db |||||

ftp.wipo.int/pub/published_pct_sequences

Sequence 900 BP; 184 A; 325 C; 253 G; 138 T; 0 U; 0 Other;

PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362999P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR P-PSDB; ABU38283.
XX

New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

Claim 14; SEQ ID NO 30023; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

CC	Bifidobacterium longum in a biological sample. A carrier containing the
CC	lactic acid bacterium Bifidobacterium longum NC22705 (NCNM 1-2618) can
CC	be used for preventing and/or treating diarrhoea brought about by pathogenic
CC	bacteria and/or rotavirus. The carrier is a food composition selected
CC	from milk, yogurt, curd, cheese, fermented milks, milk based fermented
CC	products, ice-creams, fermented cereal based products, milk based
CC	powders, infant formula, pet food or a pharmaceutical composition
CC	selected from tablets, liquid bacterial suspensions, dried oral
CC	supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC	(I) is useful in DNA arrays or chips to carry out analysis of the
CC	expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
CC	Bifidobacterium related nucleotide sequences given in the Sequence
CC	Listing from the present invention but not mentioned further within the
CC	specification. N.B. The sequence data for this patent is not represented
CC	in the printed specification but is based on sequence information
CC	supplied by the European Patent Office
XX	
SQ	Sequence 349980 BP; 70780 A; 106600 C; 104724 G; 67876 T; 0 U; 0 Other;
	Query Match 7.1%; Score 57; DB 6; Length 349980;
	Best Local Similarity 52.7%; Pred. No. 4.8e-05;
	Matches 148; Conservative 0; Mismatches 130; Indels 3; Gaps 1;
Qy	10 CCAGTGTCTCGATAATTATTCCTACCTTCATGCTCAGCGGTGACGCTGCAGACGCTGCCCTC 69
Db	230637 CGATTGGTGAGCATCATTTATTCGGGTGTATAGGTTCGAGAAGTTCTCTCACGAATTCGCTG 230696
Qy	70 GDAAGCATCTGTGGGCAGACACTACCGGGAAGTGGAAAGTGGTCTTGTGCACGGCGGTTCG 129
Db	230697 GGGTCCGGTTGTCGCGACAGCTGACGCCAATCTGGAGATTCCTGCTGGTCGACACGGCTCG 230756
Qy	130 ACCGATCGGAACCTTCGACATCGCGGAACAGTTTTCCGCCGGGAACCTCGGCTCGCGACTGGTTC 189
Db	230757 CCCGACAATATTCGCCGGCCCATGTGCCAGCGCT---GGGGCCGCCCGGACCCGGGCATACGC 230813
Qy	190 GTTCACAGCGGGCCCGATGATGTGGCCCCCTACGACGCCATGAAACGGGGCGTTCGGCGTAGCC 249
Db	230814 GTGATTCACAGCCCMAACGGGGCCCTGTCCGACGCCCGCACTCCGGTATCCGTGAGGCT 230873
Qy	250 ACAGGCGAATGGGTACTTTTTTTTAGCGGCCGACGACACCCCT 290
Db	230874 ATCGGCGGTACATATACTCTCGCCGATTCGACAGCACCCGT 230914

This Page Blank (uspto)

Qy	Qy
1	ATGACTGCGCCAGTGTCTTCGATAAATTATCGCTACCTTCAATGACGGGTGACGCTGCAA
60	60
1	ATGACTGCGCCAGTGTCTTCGATAAATTATCGCTACCTTCAATGACGGGTGACGCTGCAA
60	60

QY	61	GCCTGCCTCGAAGCATCGTCGGGCAGACCTACCGGGAAGTGGAGTGGTCTCTGTGCGAC	120
Db	61	GCCTGCCTCGAAGCATCGTCGGGCAGACCTACCGGGAAGTGGAGTGGTCTCTGTGCGAC	120
QY	121	GCGCGTTTCGACCGATCGACACCTCGACATCGGAACAGTTTCCGCCCGGAACCTCGGCTCG	180
Db	121	GCGCGTTTCGACCGATCGACACCTCGACATCGGAACAGTTTCCGCCCGGAACCTCGGCTCG	180
QY	181	CGACTGCTGTTTCACAGCGGGCCGATGATGGCCCTACGACGCGCATGAACCGCGGCTC	240
Db	181	CGACTGCTGTTTCACAGCGGGCCGATGATGGCCCTACGACGCGCATGAACCGCGGCTC	240
QY	241	GCGGTAGCCACAGCGGAATGGTACTTTTTTTAGCGCGCGACGACACCTCTACGAACCA	300
Db	241	GCGGTAGCCACAGCGGAATGGTACTTTTTTTAGCGCGCGACGACACCTCTACGAACCA	300
QY	301	ACCAGTTTGGCCAGGTAGCGCTTTCTCGCGACCATGCGGAAGCATCTTGTCTAT	360
Db	301	ACCAGTTTGGCCAGGTAGCGCTTTCTCGCGACCATGCGGAAGCATCTTGTCTAT	360
QY	361	GCGCATGTTGTATCGGTTTCGACGAAAGCCGCGCATGCCGACCTTTCGACCTCGACCGC	420
Db	361	GCGCATGTTGTATCGGTTTCGACGAAAGCCGCGCATGCCGACCTTTCGACCTCGACCGC	420
QY	421	CTCCTATTGAGACGAATTTGCGCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	480
Db	421	CTCCTATTGAGACGAATTTGCGCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	480
QY	481	GGCATCGGCCCTTACAACCTCGCTACCGAGTCTGGCGGAGCTGGAGCTTCAATATTCCG	540
Db	481	GGCATCGGCCCTTACAACCTCGCTACCGAGTCTGGCGGAGCTGGAGCTTCAATATTCCG	540
QY	541	TGCTTCTTCCAAACCCGCGCTGATTACCGCTACATGAGCGTCTGATTTCCGAATACAA	600
Db	541	TGCTTCTTCCAAACCCGCGCTGATTACCGCTACATGAGCGTCTGATTTCCGAATACAA	600
QY	601	GACATGACCGCTTACGATGAGGAGGGGAGTGAATAAGATTTCAGAAAACGGTGC	660
Db	601	GACATGACCGCTTACGATGAGGAGGGGAGTGAATAAGATTTCAGAAAACGGTGC	660
QY	661	ATGTAATCTCTGGTTGACGGTGGGAGACTTTCAGCGCGCATGCTGGCGTTTGAAGAC	720
Db	661	ATGTAATCTCTGGTTGACGGTGGGAGACTTTCAGCGCGCATGCTGGCGTTTGAAGAC	720
QY	721	AAGGAGAAATCGCGCTCGCTACCGCTTGCATGAGCGTTGATAAGCGTTTAAGCGCTT	780
Db	721	AAGGAGAAATCGCGCTCGCTACCGCTTGCATGAGCGTTGATAAGCGCTTTCCTCAA	780
QY	781	GAACGAGCGCGAACCCTAG	801
Db	781	GAACGAGCGCGAACCCTAG	801

RESULT 2	
MAP223833	MAP223833
LOCUS	Mycobacterium avium paratuberculosis gs (ba,bb,c,d) genes.
DEFINITION	4435 bp DNA linear BCT 07-JAN-1999
ACCESSION	AJ223833
VERSION	AJ223833.1 GI:3550477
KEYWORDS	gsba gene; gabb gene; gsc gene; gsd gene.
SOURCE	Mycobacterium avium subsp. paratuberculosis
ORGANISM	Mycobacterium avium subsp. paratuberculosis Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium avium complex (MAC).
REFERENCE	1
AUTHORS	Tizard,M., Bull,T., Millar,D., Doran,T., Martin,H., Sumar,N., Ford,J. and Hermon-Taylor,J.
TITLE	A low G+C content genetic island in Mycobacterium avium subsp. paratuberculosis and M. avium subsp. silvaticum with homologous genes in Mycobacterium tuberculosis
JOURNAL	Microbiology (Reading, Engl.) 144 (Pt 12), 3413-3423 (1998)

MEDLINE	99098703
PUBMED	9884234
REFERENCE	2 (bases 1 to 4435)
AUTHORS	Bull,T.
TITLE	Direct Submission
JOURNAL	Submitted (29-JAN-1998) Bull T., Surgery, St Georges Hospital Medical School, Cranmer Terrace, London, SW17 ORE, UK
FEATURES	Location/Qualifiers
source	1..4435
	/organism="Mycobacterium avium subsp. paratuberculosis"
	/mol_type="genomic DNA"
	/sub_species="paratuberculosis"
	/db_xref="taxon:1770"
	201..1232
gene	/gene="gsba"
	/gene="gsbA"
	/function="unknown"
	/codon_start=1
	/transl_table=11
	/protein_id="CA11575.1"
	/db_xref="GI:3550478"
	/db_xref="GOA:O86292"
	/db_xref="UniProt/TREMBL:O86292"
	/translation="MKRALITGITGQDGSYLAEILLKSGYEVHGLVRRASFTNTSRID HLVDPHQGARLFHYADLTGTRLVTLSSIDPDEVYNLAASHVRSFDEVHTG DTTGMGIRLEAVRLSRVDCRFYQASSEMFGASPPQNESTFFYPRSPYGAAKVFS YTTTRNYREAYGLFVANGIILFNHESPRRGETFVTRKTRAVARIRAGVQSEVYWGND AIDRWGYPBYVEGMRLQAPEDDYVLATRGYTVREFAQAADFVGLDQKHVKF DDRYLRTEVDSLVGDADRAQAQSLGWKASVHTGELARIMVDADIAASECDGTPWIDTP MLPGMGVS"
gene	1172..2191
	/gene="gsbB"
	/gene="gsbB"
	/function="unknown"
	/codon_start=1
	/transl_table=11
	/protein_id="CA11576.1"
	/db_xref="GI:3550479"
	/db_xref="UniProt/TREMBL:O86293"
	/translation="MRWHTMDRHADVWLGRSKLTTPGDLDRATPVVIAGHRLVGS ALVRRPRAEGFTNLIVRSRDEIDLTDRATFDVSETRPOVIIDAAARVGGIMANNTY PAYPLSENLAIRIITLDAAVAVRPRLLELGSSCIYKVAQPIHESALLTGPLEPTN DAFAAKIAGILQVAVRRQIGLAWISAMPTNLVYGNDFSPSGSHLLPALIRYEEA KAGAEVITWNGTTPRRELLHVDLASACFLLEHFDGNHNVNVTGVDHSISEIAD MVATAVGYIGETRWDPKDPGTPRKLLDVSAALRELGLWRPRIALKDGIDATVSWYRTNA DAVR"
gene	2467..3189
	/gene="gsc"
	2467..3189
CDS	/gene="gsc"
	/function="unknown"
	/codon_start=1
	/transl_table=11
	/protein_id="CA11577.1"
	/db_xref="GI:3550480"
	/db_xref="UniProt/TREMBL:O86294"
	/translation="MDFRLNAGLMARNVSTEMLRHFRKRLLVNQFKAYGVNVVIDVG ANSQFGSLRRAGFKSRIVSFELSPGFAQLTRESADPLWCHQYALGADETITI NVAGNAGASSSVLPMLKSHODAPPPANYIGTDVAIHRLDSVASEFLNPTDVTFLKID VQFPEKQVIAGSKSTLNESCUGMQLESLFPLYEGDMLIHEALSLVSLGRLTGLLP GFTDFRNGRLQADGIFRGDD"
gene	3335..4135
	/gene="gsd"
	3335..4135
CDS	/gene="gsd"
	/function="unknown"
	/codon_start=1
	/transl_table=11
	/protein_id="CA11578.1"
	/db_xref="GI:3550481"
	/db_xref="GOA:O88109"

/db_xref="UniProt/TREMBL:O88109"
/translation="MTAPVPSIIIPFNAAVTLQACLSIVGQTYREVEVLVDGGST
DRLTIANSFRELASRLVHSGDDPDGYDAMNRGVATGEWVFLGADDTLYEPTT
LAQVAALFVLDVGVVMSRSTKGRHAGPFDRLLLPETNLCHOSIFVRELFDD
GIQPNLYRVWADNDENIRCFSPNALLITRYMDVVISYNDMTGSMGQTDKEPRK
LPMYFWAGWETCRMLAFLEKDKENRLALRLRLRVKXVSKERSAEP"

ORIGIN		Query Match		100.0%; Score 801; DB 1; Length 4435;	
		Best Local Similarity		100.0%; Pred. No. 7.6e-160;	
		Matches 801; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	ATGACTGCGCCAGTGTTCGATAATATCCCTACCTTCAATGCGAGCGGTGACGTGCAA	60		
DB	3335	ATGACTGCGCCAGTGTTCGATAATATCCCTACCTTCAATGCGAGCGGTGACGTGCAA	3394		
QY	61	GCCTGCTCGGAAGCATCGTCGGGCGAGACTACCGGGAAGTGGTCTCTTCGAC	120		
DB	3395	GCCTGCTCGGAAGCATCGTCGGGCGAGACTACCGGGAAGTGGTCTCTTCGAC	3454		
QY	121	GGCGGTTCGACCGATCGGACCTTCGACATCGGAACAGTTTCCGCCCGGAACCTCGGCTCG	180		
DB	3455	GGCGGTTCGACCGATCGGACCTTCGACATCGGAACAGTTTCCGCCCGGAACCTCGGCTCG	3514		
QY	181	CDACTGGTCTTCAGCGGGCCCGATGATGCGCCCTACGACGCGCATGAACCGCGCGTC	240		
DB	3515	CDACTGGTCTTCAGCGGGCCCGATGATGCGCCCTACGACGCGCATGAACCGCGCGTC	3574		
QY	241	GGCGTAGCCACAGGCGAATGGGTACTTTTATGCGGCGGAGAGACACCTCTACGAACA	300		
DB	3575	GGCGTAGCCACAGGCGAATGGGTACTTTTATGCGGCGGAGAGACACCTCTACGAACA	3634		
QY	301	ACCAGCTTGGCCAGGTAGCGCTTTCTCGCGGACCATGCGGCAAGCATCTCTGTCTAT	360		
DB	3635	ACCAGCTTGGCCAGGTAGCGCTTTCTCGCGGACCATGCGGCAAGCATCTCTGTCTAT	3694		
QY	361	GGCGATGTTGATGCTTCGACGAAAGCCGCGATGCGGACCTTTTCGACCTCGACCGC	420		
DB	3695	GGCGATGTTGATGCTTCGACGAAAGCCGCGATGCGGACCTTTTCGACCTCGACCGC	3754		
QY	421	CTCCTATTGAGACCAATTTGTCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	480		
DB	3755	CTCCTATTGAGACCAATTTGTCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	3814		
QY	481	GGCATCGGCCCTTACAACTCGGCTACCGAGTCTGGGCGGACTGGGACTTCAATTCGC	540		
DB	3815	GGCATCGGCCCTTACAACTCGGCTACCGAGTCTGGGCGGACTGGGACTTCAATTCGC	3874		
QY	541	TGCTTCTCCAAACCCGGCGTGAATACCGCTACATGAGAGCTGCTGATTTCCGAATACAC	600		
DB	3875	TGCTTCTCCAAACCCGGCGTGAATACCGCTACATGAGAGCTGCTGATTTCCGAATACAC	3934		
QY	601	GACATGACCGGCTTACGATGAGGAGGAGGAGTGAAGAGTTTCAAGAAACCGGTGCA	660		
DB	3935	GACATGACCGGCTTACGATGAGGAGGAGGAGTGAAGAGTTTCAAGAAACCGGTGCA	3994		
QY	661	ATGTACTTCTGGGTTGCGAGGTGGGAGCTTCGACGCGCATGCTGGCGTTTGAAGAC	720		
DB	3995	ATGTACTTCTGGGTTGCGAGGTGGGAGCTTCGACGCGCATGCTGGCGTTTGAAGAC	4054		
QY	721	AAGGAGAAATCGCGCTCGCGCTTGGCTACGCGGTTGATAAGGGTTAAGCGCTCTCCAAA	780		
DB	4055	AAGGAGAAATCGCGCTCGCGCTTGGCTACGCGGTTGATAAGGGTTAAGCGCTCTCCAAA	4114		
QY	781	GAACGAGCGGAGAACCGTAG	801		
DB	4115	GAACGAGCGGAGAACCGTAG	4135		

RESULT 3
A63783 LOCUS
DEFINITION Sequence 4 from Patent WO9723624.
A63783 4435 bp DNA linear PAT 12-MAR-1998

ACCESSION A63783.1 GI:3717355
VERSION A63783.1
KEYWORDS
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1
AUTHORS Hermon-Taylor,J., Doran,T., Millar,D., Tizard,M., Loughlin,M.,
Sumar,N. and Ford,J.
TITLE NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA
AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY
JOURNAL Patent: WO 9723624-A 4 03-JUL-1997;
ST GEORGE S HOSPITAL MEDICAL S (GB)
COMMENT Other publication AU 1202797 19970717.
FEATURES
Location/Qualifiers
1..4435
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"

ORIGIN		Query Match		100.0%; Score 801; DB 6; Length 4435;	
		Best Local Similarity		100.0%; Pred. No. 7.6e-160;	
		Matches 801; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	ATGACTGCGCCAGTGTTCGATAATATCCCTACCTTCAATGCGAGCGGTGACGTGCAA	60		
DB	3335	ATGACTGCGCCAGTGTTCGATAATATCCCTACCTTCAATGCGAGCGGTGACGTGCAA	3394		
QY	61	GCCTGCTCGGAAGCATCGTCGGGCGAGACTACCGGGAAGTGGTCTCTTCGAC	120		
DB	3395	GCCTGCTCGGAAGCATCGTCGGGCGAGACTACCGGGAAGTGGTCTCTTCGAC	3454		
QY	121	GGCGGTTCGACCGATCGGACCTTCGACATCGGAACAGTTTCCGCCCGGAACCTCGGCTCG	180		
DB	3455	GGCGGTTCGACCGATCGGACCTTCGACATCGGAACAGTTTCCGCCCGGAACCTCGGCTCG	3514		
QY	181	CGACTGTGCTTCACAGCGGCGCGGATGATGCGCCCTACGACGCGCATGAACCGCGCGCTC	240		
DB	3515	CGACTGTGCTTCACAGCGGCGCGGATGATGCGCCCTACGACGCGCATGAACCGCGCGCTC	3574		
QY	241	GGCGTAGCCACAGGCGAATGGGTACTTTTATGCGGCGGAGAGACACCTCTACGAACA	300		
DB	3575	GGCGTAGCCACAGGCGAATGGGTACTTTTATGCGGCGGAGAGACACCTCTACGAACA	3634		
QY	301	ACCAGCTTGGCCAGGTAGCGCTTTCTCGCGGACCATGCGGCAAGCATCTCTGTCTAT	360		
DB	3635	ACCAGCTTGGCCAGGTAGCGCTTTCTCGCGGACCATGCGGCAAGCATCTCTGTCTAT	3694		
QY	361	GGCGATGTTGATGCTTCGACGAAAGCCGCGATGCGGACCTTTTCGACCTCGACCGC	420		
DB	3695	GGCGATGTTGATGCTTCGACGAAAGCCGCGATGCGGACCTTTTCGACCTCGACCGC	3754		
QY	421	CTCCTATTGAGACCAATTTGTCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	480		
DB	3755	CTCCTATTGAGACCAATTTGTCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	3814		
QY	481	GGCATCGGCCCTTACAACTCGGCTACCGAGTCTGGGCGGACTGGGACTTCAATTCGC	540		
DB	3815	GGCATCGGCCCTTACAACTCGGCTACCGAGTCTGGGCGGACTGGGACTTCAATTCGC	3874		
QY	541	TGCTTCTCCAAACCCGGCGTGAATACCGCTACATGAGAGCTGCTGATTTCCGAATACAC	600		
DB	3875	TGCTTCTCCAAACCCGGCGTGAATACCGCTACATGAGAGCTGCTGATTTCCGAATACAC	3934		
QY	601	GACATGACCGGCTTACGATGAGGAGGAGTGAAGAGTTTCAAGAAACCGGTGCA	660		
DB	3935	GACATGACCGGCTTACGATGAGGAGGAGTGAAGAGTTTCAAGAAACCGGTGCA	3994		
QY	661	ATGTACTTCTGGGTTGCGAGGTGGGAGCTTCGACGCGCATGCTGGCGTTTGAAGAC	720		
DB	3995	ATGTACTTCTGGGTTGCGAGGTGGGAGCTTCGACGCGCATGCTGGCGTTTGAAGAC	4054		

```
Qy 721 AAGGAGAAATCGCGCTTGGCTAGCGGTTGATAAGGGTTAAGCGCTCTCCAAA 780
|||||
Db 4055 AAGGAGAAATCGCGCTTGGCTAGCGGTTGATAAGGGTTAAGCGCTCTCCAAA 4114
|||||
Qy 781 GAACGAAGCGCAGAACCGTAG 801
|||||
Db 4115 GAACGAAGCGCAGAACCGTAG 4135
|||||
```

```
RESULT 4
AE017231
LOCUS
DEFINITION
Mycobacterium avium subsp. paratuberculosis str. k10, linear BCT 02-FEB-2004
16 of the complete genome.
ACCESSION
AE017231 AE016958
VERSION
AE017231.1 GI:41395599
KEYWORDS
Mycobacterium avium subsp. paratuberculosis str. k10
SOURCE
Mycobacterium avium subsp. paratuberculosis str. k10
ORGANISM
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
avium complex (MAC)
1 (bases 1 to 301068)
Li L., Bannantine, J., Zhang, Q., Amonsin, A., Alt, D. and Kapur, V.
Direct Submission
Submitted (05-SEP-2003) Biomedical Genomics Center, University of
Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108, USA
LOCATION/Qualifiers
1..301068
/organism="Mycobacterium avium subsp. paratuberculosis
str. k10"
/mol_type="genomic DNA"
/strain="k10"
/sub_species="paratuberculosis"
/db_xref="taxon:262316"
complement(198..1433)
/locus_tag="MAP1150c"
complement(198..1433)
/locus_tag="MAP1150c"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA03467.1"
/db_xref="GI:41395600"
/translation="MALDQSALEVLDRNADADRDKQAETIYQALIDAEITAVI
GAGPHERASRINQSRPKPTLSTIAGLELRIPKRSQFFPALLERRRRVDOCLF
AVVMEAYLHGTSTRKVDLVRKALGADAGISKEVSRI CADLDTVEGAFRPLSEQHF
PYVFLDQYTKARVNHVRVSVQAVVIATGVAADGRVLFQVDSGDAFWATFRLS
KTRGLSGVLVSDAHTGLRGAIEAILGASWQRCRVHFLRNVLQAVPKGSAEMVAAA
IRTIQAQPDASHVRELQDTIAGMLGRQLPKVETWLRREADDITAPADFPVLHMKIWS
TNPLRLNKEIKRRTDVGVPFNPANLRLAGSLVULVEAHQWAEKRYLSETTLLAL
HPRSDSQSVAVPAAITA"
complement(1792..2088)
/locus_tag="MAP1151c"
complement(1792..2088)
/locus_tag="MAP1151c"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA03468.1"
/db_xref="GI:41395601"
/translation="MSRPAPRSVPPAHHHEFAGRRSCPRVAETIARGYGLALRHDRP
ASPAPPGRAAAGSRFRGEQVADIPLYLGQIRHGAVNGCRKAGVPPRAAIGE"
2109..3359
/locus_tag="MAP1152"
2109..3359
/locus_tag="MAP1152"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA03469.1"
/db_xref="GI:41395602"
/translation="WDFGSLPEINSGRISYGPGSAPLAAAAAHHGLAAEMHSAAS
```

gene	3371..4735
CDS	/locus_tag="MAP1153" 3371..4735 /locus_tag="MAP1153" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AA03470.1" /db_xref="GI:41395603"
gene	4792..5145
CDS	/locus_tag="MAP1154" 4792..5145 /locus_tag="MAP1154" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AA03471.1" /db_xref="GI:41395604"
gene	5238..6200
CDS	/locus_tag="MAP1155" 5238..6200 /locus_tag="MAP1155" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AA03472.1" /db_xref="GI:41395605"
gene	6353..7747
CDS	/locus_tag="MAP1156" 6353..7747 /locus_tag="MAP1156" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AA03473.1" /db_xref="GI:41395606"
gene	7820..8524
CDS	/locus_tag="MAP1157c" complement(7820..8524) complement(7820..8524)

```
YGSIAELRTLWHGPGSTAMAAAAAFIAWLGGTAAOAEOTAAOATAAAAYDSVFAAT
VPPVIAANRALLASLIATNVLGONTTALJAATEAHYAEWMAQDRAAMVAYAGASAVAT
RLTFPGAPQSDANRAADQSAASALQLSTASSVESALSQGVSPVPAQVATATV
TAAQLPLSLTDITGILKTFNSVMGTISGPTYPLGVANLAKNWQIALUSLSPVSGTIGQ
GIGLPLHPKALTGVLPRLSLDLTGSTALSSAGTIVSASAGRAGLVLSLSPVANNWASA
VPVRTVAABELPETMLDAAPAMAVNGQQMGFPTALSSLAGRAVGGTATRAVAGSTVR
VPGAVAVDDLATTTSTVIVIPENAK"
3371..4735
/locus_tag="MAP1153"
3371..4735
/locus_tag="MAP1153"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA03470.1"
/db_xref="GI:41395603"
/translation="MVFSDFATLPPEVISTQIHVGPAGVPLLLAAAAWDGLAELHGT
AASVASYISELVESWQSSGMAAAAPYAWLSATAGLAEQTAQAQAAVAAAYEAA
AVAATVPVVAANRSLATLLOTNILGONDAATAATEDEYGMWAQDVAAMFSYAAA
SESAGELTPEQAPATTNDAGPATQAAAAAESTSLSSWLKOLEAVITSLTGQYTOP
WQNLIGATGSTETAPFWETIYSSISQVGTQATVNVNATIGLVQWQKNNFFIYAPW
SAAMAKSLGALASPGHIAAKPVSAGLVNAPTVGKLSVPAGWATAATAPALRLAS
TALPATSLAAAAATDFGLNEATLGLAGGALGSPARVVSTTGTIETPATITGRRQG
PVKLDRIIAQLQEMPEQVQHNVNDEAGLDDLVARLRTTPGVHAVHVITDGEQVAVGSPA
SKLG"
4792..5145
/locus_tag="MAP1154"
4792..5145
/locus_tag="MAP1154"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA03471.1"
/db_xref="GI:41395604"
/translation="MKLLALLGVALGIVTAVPAQAVPEGEDEAATDDNNEVFIADLHK
VGISFQDPGQAVSAGKAVCVGLLARGVSLQLLNDLRNPNALTITNGAQPATISAKSY
CPROLESAGKGGG"
5238..6200
/locus_tag="MAP1155"
5238..6200
/locus_tag="MAP1155"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA03472.1"
/db_xref="GI:41395605"
/translation="MDFGTRAPESNGRHSHPGAGSLMDAAKAWGLQALADALT
VYAVPIAANRLRRIISLAENCLGQLSPAVADADADYERMAQDRAAMVAYARASAGV
WKLTFPDSPPADPAAGVTPSPRNWALTAPEVITTGROVISTIDALGALSVSPL
ATFTLSLSTAGLSRLSSLSAPLDFALNHLNALKAAALGRAAVMCSGPRFGGGGTN
SAALRVSVGGAAGIILSVPRATRSVPSPNSPAPLDRSQ"
6353..7747
/locus_tag="MAP1156"
6353..7747
/locus_tag="MAP1156"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA03473.1"
/db_xref="GI:41395606"
/translation="MKRLSSVDAAFNSAETAGTWHMVYCALAICDPKSAEYSFELRQ
LLIBRLPEVQLRWVRWVGAPLGLDRPMFVEDEELDVPDHLRRIIGVPAPGGRREIELV
GRMSYKLDRLSRKLEWMMVTEGLDGNAGGRFATLTKMHHAIVDGVSGAGLGEILLDAT
PEPRPQOETVGSILVGFQPLGLERRAVGALINVGKTPFRIARLEQTVRQOQVATLV
RSRPRYPFAPKTRFNAPVSPHRRVTGCRVELARAKAIKDAYDKLVKNDVVLALVAGAV
RDVLQKRGELPAKPLIAQIPVSTRTDENKDDIGNKVSMTVSLATHIEDPCRLEAIH
DSTQSAKLMAKALSQAQIMGLTETTPGLLHAAARVATASGLSENLAFINLVSNVPG
PPFPLMAGAKLDSIGPMDVALNITCFSTYIDYDFGVFTTPEVANDIDMDADRIEP
ALTDLEKAAAGETG"
complement(7820..8524)
/locus_tag="MAP1157c"
complement(7820..8524)
```

gene
/locus_tag="MAP1157c"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA03474.1"
/db_xref="GI:41395607"
/translation="MAVVASRRHWPASAAVHGARWVDTAPLEILHSNRNPLPLGR
IHSRLSEVALIDGVPTTPARTAVDLGWHVPVDAVAVIIDLRLATGCAVACOL
LADYVPRGRGWSARAALGLADAGQSPKETWLRLLRAGLPRPOTQIPVSEFFGV
TYLDMGHEDEWAAAYEGEQHRRDRWQYTWDIRRETLERLGIWIVQVADRPADI
VSRVSAALARGVTLA"
complement(8842..10104)
/gene="lipo"
/locus_tag="MAP1158c"
complement(8842..10104)
/gene="lipo"
/locus_tag="MAP1158c"
/codon_start=1
/transl_table=11
/product="lipo"
/protein_id="AA03475.1"
/db_xref="GI:41395608"
/translation="MRTKRPRLARATLELANANGFRPLARKGYSTVLVFWGWA
SEVPGIYFSASLLDLRGRGDFAGRRGKAALATAASWAILGVKTRIGTITGAVPL
EAGLDQDDYTEALNKLPQSRPTRSGRRLPLGNIVARRRYVEKTVNVSYGPHRA
NLADIRWRDLPRDGKAPVLLQVPGGAWAIGMRPQAYPLMSHLAARGWCVSIGYR
SPRHTWPDHIVDVKRALAKVENIARYGDPNFVAITGSGAGHLCSLAALTDPNDKY
QPGREDATSVVAAVPVYGVYDWFTEGEGRRFVQLLEKFVVKKKEATHRDIYVDAS
PIRLRLADAPPFVHLGRDSDLI PVGEAQEFVEELRAVSKSPVAYAEPLPHQAHDIF
SSPRHRAEAVAFSLVWYATNPPRD"
complement(10113..11732)
/gene="fad12"
/locus_tag="MAP1159c"
complement(10113..11732)
/gene="fad12"
/locus_tag="MAP1159c"
/codon_start=1
/transl_table=11
/product="FadD12"
/protein_id="AA03476.1"
/db_xref="GI:41395609"
/translation="MNFNVRTELGLIATMRRLIAPMRPDYRLIRIAAMRREGMGMT
SGFAAQRDPRLDVERGLSTWQDLRCDALAAALQALQSGAPAVIGIMCRNHR
GFVEALAAQRDGLADIVLLNTS FAPALADVTREGVNAVITYBEFTATVDALAGRP
DAIRIVAMTDHOHTVDKLTASKAGARPITRGKGMILLSTGTTPKAKOSGNN
AGIGTLKALIDRTPRAEPPVIVAPMFHAFSGQLLILAAAFACPVILTRKEDPEALT
DLIDHRATGLVVPVMDRLMDLPAEVRRYECSRLPAAASGRMRPDVVAFMDE
FGDVYNNYNATEAGMIATATPADLRAAPDTAGRPAGGTEIRILDPFNEFNPAGEVGT
IYVRNTOFDYTGSSKDFHEGMSGDGLYDSAGRLFVVRDDEMIIVSGENVYP
IEVEKTLATPDVAEAAVIGVDDEQYQRLAAFFVVLPAPEARTTPEALKQHVDRNLAY
KVPREISVLDELPRSTGKILRADLRVGG"
complement(11736..12545)
/locus_tag="MAP1160c"
complement(11736..12545)
/locus_tag="MAP1160c"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA03477.1"
/db_xref="GI:41395610"
/translation="MTDSNADGGDIGKDFPVLQRLMTVMRVPVLKAYFRSEVHGLES
FPFGGALVVGHSNGMFPMDVPIFSHFYDKYDPRPIYTLSDILMTGPTGDPFRRT
GYIRANRENAALALRSGGWIVFPGGDYDAVRPTLAENVIDFGKRGVRTAIEAGVP
IVPAVSGSQETQLVLTGTWLAEELGIKRLIRSNAILPVSGFPFGLSAVPPNLP
AKIVTRVLDPIADIAQFSDVDVEDEYVRSVMQALNELAERFPILG"
Query Match
Best Local Similarity 100.0%; Score 801; DB 1; Length 301068;
Matches 801; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 GCCTGCTCGGAAGCATCGTGGGCGAGACCTACCGGGAAGTGGTGGTCTTGTGCGAC 120
DB 91304 GCCTGCTCGGAAGCATCGTGGGCGAGACCTACCGGGAAGTGGTGGTCTTGTGCGAC 91363
QY 121 GCGGCTTCGACCGCATCGGACCTCTGCAATCGGGAACAGTTCCTCGCCCGGAACTCGGCTCG 180
DB 91364 GCGGCTTCGACCGCATCGGACCTCTGCAATCGGGAACAGTTCCTCGCCCGGAACTCGGCTCG 91423
QY 181 CGACTGGTTCGTTCAAGCGGGGCGCATGATGCGCCCTACGAGCGCATGAACCGCGCGCTC 240
DB 91424 CGACTGGTTCGTTCAAGCGGGGCGCATGATGCGCCCTACGAGCGCATGAACCGCGCGCTC 91483
QY 241 GCGGTAGCCACAGCGCAATGGGTACTTTTTTATAGCGCGCGACGACACCTCTTACGAAACA 300
DB 91484 GCGGTAGCCACAGCGCAATGGGTACTTTTTTATAGCGCGCGACGACACCTCTTACGAAACA 91543
QY 301 ACCAGTTGGCCAGGTAGCGCTTTCTCGGCGACCATGCGGCAAGCATCTTGTCTAT 360
DB 91544 ACCAGTTGGCCAGGTAGCGCTTTCTCGGCGACCATGCGGCAAGCATCTTGTCTAT 91603
QY 361 GCGCATGTTGTGATGCGTTTCCAGCAAGCGCGCATGCGGACCTTTTCGACCTCGACCGC 420
DB 91604 GCGCATGTTGTGATGCGTTTCCAGCAAGCGCGCATGCGGACCTTTTCGACCTCGACCGC 91663
QY 421 CTCTATTGTGAGCGAATTTTGTGCCACCAATCGATCTTTTACCGCGCTGAGCTTTTCGAC 480
DB 91664 CTCTATTGTGAGCGAATTTTGTGCCACCAATCGATCTTTTACCGCGCTGAGCTTTTCGAC 91723
QY 481 GGCATTCGCGCTTACAACCTCGCTACCGAGTCTGGGCGGACTGCGGACTTCAATATTCG 540
DB 91724 GGCATTCGCGCTTACAACCTCGCTACCGAGTCTGGGCGGACTGCGGACTTCAATATTCG 91783
QY 541 TGCTTCTCCAAACCGCGCGCTGATACCGCTTACATGAGACGCTGATTTCCGAATACAAC 600
DB 91784 TGCTTCTCCAAACCGCGCGCTGATACCGCTTACATGAGACGCTGATTTCCGAATACAAC 91843
QY 601 GACATGACCGGCTTTCAGCATGAGCGGGAAGTCTGATAAGAGTTCAGAAAAACGCTGCCA 660
DB 91844 GACATGACCGGCTTTCAGCATGAGCGGGAAGTCTGATAAGAGTTCAGAAAAACGCTGCCA 91903
QY 661 ATGTACTTCTGGTTCAGAGGTGGAGACTTTCGAGGCGCATGCTGGCGTTTGAAGAAC 720
DB 91904 ATGTACTTCTGGTTCAGAGGTGGAGACTTTCGAGGCGCATGCTGGCGTTTGAAGAAC 91963
QY 721 AAGGAGATCGCCTCTGGCCTTCGTAAGCGGTTGATAAGGTTAAGCGCTTCCCAA 780
DB 91964 AAGGAGATCGCCTCTGGCCTTCGTAAGCGGTTGATAAGGTTAAGCGCTTCCCAA 92023
QY 781 GAACGAAGCGCAGAACCGTAG 801
DB 92024 GAACGAAGCGCAGAACCGTAG 92044
RESULT 5
LOCUS A63800 801 bp DNA linear PAT 12-MAR-1998
DEFINITION Sequence 21 from Patent WO9723624.
ACCESSION A63800
VERSION A63800.1 GI:3717372
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1
AUTHORS Hermon-Taylor,J., Doran,T., Millar,D., Tizard,M., Loughlin,M.,
Sumar,N. and Ford,J.
TITLE NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA
AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY
JOURNAL Patent: WO 9723624-A 21 03-JUL-1997;
COMMENT ST GEORGE S HOSPITAL MEDICAL S (GB)
FEATURES Other publication AU 1202797 19970717.
Location/Qualifiers

```
source
1. .801
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
CDS
1. .801
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA03748.1"
/db_xref="GI:3717373"
/translation="MTAPVFSIIIPTFNAVTLQKLSIVGQTYREVVLVPGGST
DRTLDIANSFPELGRSLVHSGDDGPDYDMNRGVGATGEWVLFGLADTLYEPTT
LAQVAAFHLVGDVVMRSTKSRHAGPFDLDRLLFETNLCHOSIYRRELF
GIGPNLYRYWADENIRCFSPALI TRYMDVVISYNDMTGFSMEQGTDKFERK
LPMYFWAGNFCRMLAPLKDKNRRLALRFLIRVKVKSERSAEP"

ORIGIN
Query Match          99.8%; Score 799.4; DB 6; Length 801;
Best Local Similarity 99.9%; Pred. No. 1.6e-159;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGACTGGCCAGTGTTCGATAATATTCCTCAATGCGAGCGGTGACGCTGCAA 60
Db 1 ATGACTGGCCAGTGTTCGATAATATTCCTCAATGCGAGCGGTGACGCTGCAA 60
QY 61 GCCTGCCCTCGGAAGCATCGTCGGGCAGACCTACCGGGAAGTGGAAGTGGTCTTGTGAC 120
Db 61 GCCTGCCCTCGGAAGCATCGTCGGGCAGACCTACCGGGAAGTGGAAGTGGTCTTGTGAC 120
QY 121 GCGCGTTCGACCGATCGGACCCCTGCACATCGGAAACAGTTTCGCGCCCGGAACCTCGGCTCG 180
Db 121 GCGCGTTCGACCGATCGGACCCCTGCACATCGGAAACAGTTTCGCGCCCGGAACCTCGGCTCG 180
QY 181 CGACTGGTGTTCACAGCGGGGCCGATGATGTCGCCCTACGAGCGCATGAAACGCGGGCGTC 240
Db 181 CGACTGGTGTTCACAGCGGGGCCGATGATGTCGCCCTACGAGCGCATGAAACGCGGGCGTC 240
QY 241 GCGGTAGCCACAGGCGAATGGGTACTTTTTTTAGCGCGCGACGACACACCTCTACGAACA 300
Db 241 GCGGTAGCCACAGGCGAATGGGTACTTTTTTTAGCGCGCGACGACACACCTCTACGAACA 300
QY 301 ACCAGTTTGGCCAGGTAGCGCTTTTCTCGCGCACCATGCGGCGAAGCCATCTTGTCTAT 360
Db 301 ACCAGTTTGGCCAGGTAGCGCTTTTCTCGCGCACCATGCGGCGAAGCCATCTTGTCTAT 360
QY 361 GCGGATGTTCGATCGGTTCGACGAAAGCCGGGATGCGGACCTTTTCGACCTCGACCGC 420
Db 361 GCGGATGTTCGATCGGTTCGACGAAAGCCGGCATGCGGACCTTTTCGACCTCGACCGC 420
QY 421 CTCCTATTGAGACGAAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
Db 421 CTCCTATTGAGACGAAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
QY 481 GGCATCGGCCCTTACAACCTCGGCTACCGAGTCTCGGGCGGACTGGGACTTCAATATTCGC 540
Db 481 GGCATCGGCCCTTACAACCTCGGCTACCGAGTCTCGGGCGGACTGGGACTTCAATATTCGC 540
QY 541 TGCCTTCCAAACCGCGCGTGAATACCGCTACATGAGCGTGTGATTTCCGAATACAAAC 600
Db 541 TGCCTTCCAAACCGCGCGTGAATACCGCTACATGAGCGTGTGATTTCCGAATACAAAC 600
QY 601 GACATGACCGGCTTCAGCATGAGGCGAGGACTGTATAAGAGTTCAGAAAACCGCTGCCA 660
Db 601 GACATGACCGGCTTCAGCATGAGGCGAGGACTGTATAAGAGTTCAGAAAACCGCTGCCA 660
QY 661 ATGTACTTCTGGGTTGCGAGGTGGAGACTTCGAGCGCATGCTGCGCGTTTTTGAAGAC 720
Db 661 ATGTACTTCTGGGTTGCGAGGTGGAGACTTCGAGCGCATGCTGCGCGTTTTTGAAGAC 720
QY 721 AAGGAGATCGCGCTCTGCGCTTGGGTACGCGGTTGATAAGGGTTAAGCGCTCTCCAAA 780
Db 721 AAGGAGATCGCGCTCTGCGCTTGGGTACGCGGTTGATAAGGGTTAAGCGCTCTCCAAA 780
QY 781 GAACGAAGCGCAGAACCGCTAG 801
```

```
Db 781 GAACGAAGCGCAGAACCGTAG 801
|||||
RESULT 6
LOCUS A63782 7995 bp DNA linear PAT 12-MAR-1998
DEFINITION Sequence 3 from Patent WO9723624.
ACCESSION A63782
VERSION A63782.1 GI:3717354
KEYWORDS .
SOURCE unidentified
ORGANISM unidentified
REFERENCE unclassified.
1.
Hermion-Taylor,J., Doran,T., Millar,D., Tizard,M., Loughlin,M.,
Sumar,N. and Ford,J.
NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA
AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY
Patent: WO 9723624-A 3 03-JUL-1997;
ST GEORGE S HOSPITAL MEDICAL S (GB)
Other publication AU 1202797 19970717.
Location/Qualifiers
source
1. .7995
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"

ORIGIN
Query Match          99.8%; Score 799.4; DB 6; Length 7995;
Best Local Similarity 99.9%; Pred. No. 1.7e-159;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGACTGGCCAGTGTTCGATAATATTCCTCAATGCGAGCGGTGACGCTGCAA 60
Db 4947 ATGACTGGCCAGTGTTCGATAATATTCCTCAATGCGAGCGGTGACGCTGCAA 5006
QY 61 GCCTGCCCTCGGAAGCATCGTCGGGCAGACCTACCGGGAAGTGGAAGTGGTCTTGTGAC 120
Db 5007 GCCTGCCCTCGGAAGCATCGTCGGGCAGACCTACCGGGAAGTGGAAGTGGTCTTGTGAC 5066
QY 121 GCGCGTTCGACCGATCGGACCCCTGCACATCGCAACAGTTTTCGCGCCCGGAACCTCGGCTCG 180
Db 5067 GCGCGTTCGACCGATCGGACCCCTGCACATCGCAACAGTTTTCGCGCCCGGAACCTCGGCTCG 5126
QY 181 CGACTGGTGTTCACAGCGGGGCCGATGATGTCGCCCTACGAGCGCATGAAACGCGGGCGTC 240
Db 5127 CGACTGGTGTTCACAGCGGGGCCGATGATGTCGCCCTACGAGCGCATGAAACGCGGGCGTC 5186
QY 241 GCGGTAGCCACAGGCGAATGGGTACTTTTTTTAGCGCGCGACGACACACCTCTACGAACA 300
Db 5187 GCGGTAGCCACAGGCGAATGGGTACTTTTTTTAGCGCGCGACGACACACCTCTACGAACA 5246
QY 301 ACCAGTTTGGCCAGGTAGCGCTTTTCTCGCGCACCATGCGGCGAAGCCATCTTGTCTAT 360
Db 5247 ACCAGTTTGGCCAGGTAGCGCTTTTCTCGCGCACCATGCGGCGAAGCCATCTTGTCTAT 5306
QY 361 GCGGATGTTCGATCGGTTCGACGAAAGCCGGGATGCGGACCTTTTCGACCTCGACCGC 420
Db 5307 GCGGATGTTCGATCGGTTCGACGAAAGCCGGGATGCGGACCTTTTCGACCTCGACCGC 5366
QY 421 CTCCTATTGAGACGAAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
Db 5367 CTCCTATTGAGACGAAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 5426
QY 481 GGCATCGGCCCTTACAACCTCGGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTCGC 540
Db 5427 GGCATCGGCCCTTACAACCTCGGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTCGC 5486
QY 541 TGCCTTCCAAACCGCGCGCTGATTAACCGCTACATGAGCGTGTGATTTCCGAATACAAAC 600
Db 5487 TGCCTTCCAAACCGCGCGCTGATTAACCGCTACATGAGCGTGTGATTTCCGAATACAAAC 5546
```


ORIGIN		/db_xref="GI:3550476" /db_xref="GOA:O88109" /db_xref="UniProt/TREMBL:O88109" translation="MTAPVFSIIITFNAAVTLQACIGSIVGQTYREVVVVDGGST LRTDIIANSFPELGSRLVHSGDDGDPYDMNRGVGATGEWLVFLGADDTLYEPTT LAQVAALFGDHAASHLVYGDVVMRSTKSHAGPFDLRLLPETNLCHQSIFPYRELF GIGPNLYRYWADWDFNIRCFSPALITRYMDVVISYNDMTGFSMRQGTDEKFRKR LPMYFWAGWETCRRLAFLKOKENRLALTRLIRVKAVSKERSAEP"			
Query Match	99.8%;	Score 799.4;	DB 1;	Length 8938;	
Best Local Similarity	99.9%;	Pred. No. 1.7e-159;			
Matches 800;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;	
QY	1	ATGACTGGCCAGTGTTCGATAAATTATCCCTCAATGACGCGTGACGCTGCAA	60		
Db	5890	ATGACTGGCCAGTGTTCGATAAATTATCCCTCAATGACGCGTGACGCTGCAA	5949		
QY	61	GCTGCTCGGAAGCATCGTGGGCAGACCTACGGGAAGTGAAGTGTCTTTGTTCGAC	120		
Db	5950	GCTGCTCGGAAGCATCGTGGGCAGACCTACGGGAAGTGAAGTGTCTTTGTTCGAC	6009		
QY	121	GCGGTTTCGACCGATCGGACCTCGACATCGCAACAGTTCCTCGCCCGGAACCTCGGCTCG	180		
Db	6010	GCGGTTTCGACCGATCGGACCTCGACATCGCAACAGTTCCTCGCCCGGAACCTCGGCTCG	6069		
QY	181	CGACTGCTGTTTCAACGCGGGCCGATGATGGCCCTTACGACGCCATGAACCGCGGCTC	240		
Db	6070	CGACTGCTGTTTCAACGCGGGCCGATGATGGCCCTTACGACGCCATGAACCGCGGCTC	6129		
QY	241	GCGGTAGCCACAGCGCAATGGGTACTTTTTTAGCGCGCGACGACACCTCTACGAACTCA	300		
Db	6130	GCGGTAGCCACAGCGCAATGGGTACTTTTTTAGCGCGCGACGACACCTCTACGAACTCA	6189		
QY	301	ACCAGTTGGCCAGGTAGCGCTTTTCTCGCGGACCATCGGCAAGCCATCTTTGTCTAT	360		
Db	6190	ACCAGTTGGCCAGGTAGCGCTTTTCTCGCGGACCATCGGCAAGCCATCTTTGTCTAT	6249		
QY	361	GCGGATGTTGTGATCGTTTCGACGAAAGCCGGCATGCGGACCTTCGACCTCGACCGC	420		
Db	6250	GCGGATGTTGTGATCGTTTCGACGAAAGCCGGCATGCGGACCTTCGACCTCGACCGC	6309		
QY	421	CTCCTATTTGAGACGAATTTGCGACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	480		
Db	6310	CTCCTATTTGAGACGAATTTGCGACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	6369		
QY	481	GGCATCGGCCCTTACAACCTCGCTACCGAGTCTGGCGGACCTGGGACTTCAATATTCGC	540		
Db	6370	GGCATCGGCCCTTACAACCTCGCTACCGAGTCTGGCGGACCTGGGACTTCAATATTCGC	6429		
QY	541	TGCTTCTCCAAACCCGGCGTGAATACCGCTACATGGACGCTCTGATTTCCGGAATACAAC	600		
Db	6430	TGCTTCTCCAAACCCGGCGTGAATACCGCTACATGGACGCTCTGATTTCCGGAATACAAC	6489		
QY	601	GACATGACCGGCTTCAGCATGAGCGAGGACCTGATAAGAGTTTCAGAAACCGGCTGCCA	660		
Db	6490	GACATGACCGGCTTCAGCATGAGCGAGGACCTGATAAGAGTTTCAGAAACCGGCTGCCA	6549		
QY	661	ATGTACTTCTGGGTTGCAAGGTGGAGACTTGCAGGCGCATGCTGGCGTTTTTGAAGAC	720		
Db	6550	ATGTACTTCTGGGTTGCAAGGTGGAGACTTGCAGGCGCATGCTGGCGTTTTTGAAGAC	6609		
QY	721	AAGGAGATCGCGCTTCGCTTCGTAGCGGTTGATAAGGTTAAGGCGCTCTCCAAA	780		
Db	6610	AAGGAGATCGCGCTTCGCTTCGTAGCGGTTGATAAGGTTAAGGCGCTCTCCAAA	6669		
QY	781	GAACGAAGCGCAGAACCGTAG	801		
Db	6670	GAACGAAGCGCAGAACCGTAG	6690		

LOCUS	AF125999	25861 bp	DNA	linear	BCT 02-OCT-2003
DEFINITION	Mycobacterium avium glycopeptidolipid biosynthesis and daunorubicin resistance gene clusters, complete sequence; and transport membrane protein gene, partial cds.				
ACCESSION	AF125999				
VERSION	AF125999.1				
KEYWORDS	GI:4416460				
SOURCE	Mycobacterium avium				
ORGANISM	Mycobacterium avium Bacteria; Actinobacteria; Actinobacteridae; Actinobacteriales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium avium complex (MAC) 1 (bases 1 to 25861) Eckstein,T.M., Belisle,J.T. and Inamine,J.M. Proposed pathway for the biosynthesis of serovar-specific glycopeptidolipids in Mycobacterium avium serovar 2 Microbiology (Reading, Engl.) 149 (Pt 10), 2797-2807 (2003) 14523113 2 (bases 1 to 25861) Eckstein,T.M., Brennan,P.J., Inamine,J.M. and Belisle,J.T. Identification of gene cluster involved in glycopeptidolipid biosynthesis and of a gene cluster encoding daunorubicin resistance in two strains of Mycobacterium avium serovar 2 Unpublished 3 (bases 1 to 25861) Eckstein,T.M., Brennan,P.J., Inamine,J.M. and Belisle,J.T. Direct Submission Submitted (04-FEB-1999) Microbiology, Colorado State University, Fort Collins, CO 80523-1677, USA Location/Qualifiers 1. .25861 /organism="Mycobacterium avium" /mol_type="genomic DNA" /strain="TMC 724" /db_xref="taxon:1764" 461. .1597 /gene="mtfA" 461. .1597 /gene="mtfA" /note="putative; similar to methyltransferases mtfB and mtfC in this gene cluster" /codon_start=1 /transl_table=11 /product="methyltransferase mtfA" /protein_id="AAD20361.1" /db_xref="GI:4416461" /translation="MNRWALSTDGQWFLKLRPADTRIHMLKWLKAVLLGALLLAA CYFVPLAVGIGAILLLLCARIPGRDRYIPNLRYARDRIYDDQYREFRRTLAEL RRRIIGHTLLWEASOLPQCAENSELLDLGVIGWSTRLIPDTCRTVYGFDTFS GLVEDRLEDRIVKRGAFSLSEPAQRFIDTGVITINDGVPAALGRDVRKIGSTYD TLAPLADRPAAPIRLPHMDLDTVESCLHALETCKDHVVGSIILVDEYLVNCEMRA FYDQKRYELQWYRWANGLEMENNVEMTSRWKRWLYSIAAIFGYLLGLDGRFLWAC FREPFWFLNAPAEIDIFFILGAAGSRKSVSIETIGLGLKLAAPH" 1864. .2685 /gene="mtfB" 1864. .2685 /gene="mtfB" /note="putative; similar to methyltransferase mtfC of this gene cluster and to Micromonospora griseorubida mycinamin III-O-methyltransferase, PIR Accession Number S51595" /codon_start=1 /transl_table=11 /product="methyltransferase mtfB" /protein_id="AAD20362.1" /db_xref="GI:4416462" /translation="MTDRMRSPFLDTRSAAYLDLRLRLTRYGSDLELPVGMNLYGRP LFSTRKLMVLRKPRFNHARDLWDPADALTIMGORLTSLQVCVETIKDLIPGDL VECGWRGGASILMRVLSAVGDEKRCWLCDSPGVPPDPDTHYQADKGLKRAAG ILAVPEAQVRANFERYGLDDRVPFVPGFKDTLDQAPIERIALVRLDGLDYESTIQ LDALYPRLSAGGICIIDYHDAIDCRQAVTDYRSEHGVTAPIEDGTGVLWRKS" 2807. .4084 /gene="gtfA" 2807. .4084				

```

/gene="gtfa"
/notes="putative; similar to rhamnosyltransferase rtfa and
glycosyltransferase gtfb in this gene cluster"
/codon_start=1
/transl_table=11
/product="glycosyltransferase gtfa"
/protein_id="AAD20363.1"
/db_xref="GI:4416463"
/translation="MKCVLASVYGRGDIPESSVVAARELQRRGHDMIMAVPPDSVFTF
AAGLTAVPYGLDSQAWLDVYRNFMTSPFRGFWKMGEMRMWMDLSDQCWAQNMNTT
LISLADGSLDPAGOSYOEPAAVAEYVDLPATLHVPMRPMNGOVSVLPLAPLARS
MTAPDFCWRLNKVEDQRRRLGLPKATGSPRIARGLSLEIQAYDEVCPGLAE
WANWQGRPFVGLTWELTREDDEVMRWIRGTPPICFGSGMPVESPADVDMIGT
ACAQGERALLCAGSDSGVPADPKVGVPNVYATTFPACRAVVRHGGSGTTAASL
RAGVPTLILSDANQITWGAQIKRLKVGATRFSTATDRDSLVDLRLKILAPDYARRR
EIAARWTKPADSVKMAADVEDFVQSRSTGRS"
4488..5774
/gene="rtfa"
/notes="rtfa"
/notes="similar to glycosyltransferases gtfa and gtfb in
this gene cluster"
/codon_start=1
/transl_table=11
/product="rhamnosyltransferase rtfa"
/protein_id="AAD20364.1"
/db_xref="GI:4416464"
/translation="MKFAVASVYGRGDIPECLAIQRELMRGHSVRMAVPPNVLVDLAE
AAGLSAIAYGDMHDPFMDERTDFPKNPLRGPTVLIAREMPEVLKNQEMSSALMSV
GAGADLLFTQYLQDLAINADYDYGIPWATLHLYPMRNGQLLPVPLPGLTGMVY
DMWCWMKKAEDRRRLGLPTATVAPRIARGLSLEIQAYHEVCPGLAEWAKW
GQRPFLVGLTWELTREDDEVLISVIGQTPPICFGSGMPVESPAETIIGTACA
ELGERALI CSWGSDFSNVPHADHKVGVNYTKIPACRAVVRHGGSGTTAAGLRAG
IPSLTLWTAQDPLWGSRIKOLKVCTSRFSATTPETLIVADLRLKILAPYLTRARELS
SRMSKPASAGRAVDLLEPARGSCIPRAEQRR"
5876..5876
/gene="mtfc"
/notes="mtfc"
/notes="mtfc"
/notes="putative; similar to methyltransferase mtfb in this
gene cluster and to Micromonospora griseorubida mycinamin
III-O-methyltransferase, PIR Accession Number S51595"
/codon_start=1
/transl_table=11
/product="methyltransferase mtfc"
/protein_id="AAD20365.1"
/db_xref="GI:4416465"
/translation="MTDHDTRFAYLDLRLDLTRYGSDLPVGLYRLGRPLFSTRNL
MLVRKRPNKRRARDLGLDMPADALTMIGMKRLTSLQHCVTLEEDVPGLVFCGVWR
GGASILMRAVLAAYGDEKRCVWLCDSPAGVPPDPTVNYKADKGIHLRHRHAILGVPLE
NVKANFRYGLLDDQVRFPVGMFKDLDKADPDRISVLRLDGLDYESTIQALDALYPR
LSPGGFCIVDDVYHAKACAQAVTDYRTQGTAEIVEIDGTGVLWRKP"
6776..7594
/gene="mtfd"
/notes="mtfd"
/notes="putative; similar to methyltransferases mtfb and
mtfc in this gene cluster"
/codon_start=1
/transl_table=11
/product="methyltransferase mtfd"
/protein_id="AAD20366.1"
/db_xref="GI:4416466"
/translation="MHADVMVAATVGLTPTGLRMWVRTEYWLARTLLPDVYANDA
LITFNHFADLDPDFQRYRGRSLKEDWYQWVRHVHVLNRAAASASLSDGDFVE
CGVSGVJUSSIMEDRLRGKTYLLDFTAGIDPFREVTNRAAGALEISKGLHRL
GMYRANGDVGFANFYKXNRIIVGAVPETLDQVDPASVYLIHDMNCAPPEVAALRF
FWPRLTPGAFVLLDDYANRGRDEQRAMDEVAABFGVKICTLPTGQGLIKRPR"
7668..8448
/gene="dhga"
/notes="dhga"
/notes="putative; similar to the Streptomyces coelicolor
```

```

product encoded by the sequence presented in GenBank
Accession Number AL031350 and to Mycobacterium
tuberculosis Rv1544 and Rv2509"
/codon_start=1
/transl_table=11
/product="dehydrogenase dhga"
/protein_id="AAD20367.1"
/db_xref="GI:4416467"
/translation="MALAMITGASSIGLELAELFAORGYDLVVAENDGIYPASDSL
SRMGVDPVQVDLRTPDGVEKLVREATEGGRGLDAAALNAGVGGSGFVEGELADL
SIIDLNVSTHLAKLVLNRNTRNTKVLTSLSLASMAGPPEAVYHASKSFVQSFA
GHRKDELDTGITVLSLMPGTETNFPFRAGNQTRAGCKPKDDPAEVARQCFEAMMR
GHRKVGSLTKTKTVGVALRFLPDSWKAASRVNSTPAGRR"
complement(8454..9710)
/gene="gtfb"
complement(8454..9710)
/gene="gtfb"
/notes="putative; similar to glycosyltransferase gtfa and
rhamnosyltransferase rtfa in this gene cluster"
/codon_start=1
/transl_table=11
/product="glycosyltransferase gtfb"
/protein_id="AAD20368.1"
/db_xref="GI:4416468"
/translation="MRFALASYGTRGDIPESSAAVGRELLRGRHVDLRAVPPBLVGFVE
STGLAAYPGKQVQFDEEFRLNNWRDFRNIRLVKVMDDPIIKYWADASATLKS
AEGADLLSTGLNPFQAAANVAEYDIPLASLHHPMRANGOLVPMNPSPVLRSTWAAI
BWLFWRSKTKVDNAQRQLGPKATRSQRRIASHGLWLEIQAYDEVCPGLAEWAHW
GDRPFGVGLTWELTREDDDVASWIAAGTTPICFGSGSIPLESATVEMIGAACAR
LGERALICFGDITVGRVDFHVGVPNVYAVFPACRAVVRHGGSGTTAASLRAGI
PTLMASSIEDQYWAQIKRLKVGATRFSTATTSETLVADLRLTILAPDYATRADLAA
RMTKPASISIEPTADLBEARRKTLG"
9852..10505
/notes="low gc"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD20369.1"
/db_xref="GI:4416469"
/translation="MSRIGNSLVITIALNSARNKSLPTRLAAAYGWLGDHLVSPPA
IAKORHLNVFKSNRHRI FVEAGTYGQTTAFFIPHADQVSVLHDLGLFAAQRFA
QQPNVTLVRGSLVEIPRIVATSSPPLVDLGHFSGTAGRGEEMEAESTLGRAD
VTPAGTTIVIDDLRLFGSLGSPFPQLDTITTSARAAFPFGLIRTGLDSIVVEVPLGN"
complement(10567..11523)
/gene="hlpa"
complement(10567..11523)
/gene="hlpa"
/notes="putative; similar to Prevotella intermedia product
encoded by the sequence presented in GenBank Accession
Number AF052516"
/codon_start=1
/transl_table=11
/product="hemolytic protein hlpA"
/protein_id="AAD20370.1"
Query Match 99.8%; Score 799.4; DB 1; Length 25861;
Best Local Similarity 99.9%; Pred. No. 1.7e-159;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGACTGCGCCAGTGTCTTCGATAATTATTCCTTCAATGACGCGGTGACGCTGCAA 60
Db 17522 ATGACTGCGCCAGTGTCTTCGATAATTATTCCTTCAATGACGCGGTGACGCTGCAA 17581
Qy 61 GCCTGCTCGGAAGCATGTCGGCGACGCTACCGGGAGTGGAGTGGTCTCTTCTCGAC 120
Db 17582 GCCTGCTCGGAAGCATGTCGGCGACGCTACCGGGAGTGGAGTGGTCTCTTCTCGAC 17641
Qy 121 GGCGGTTTCGACCGCATCGGACCTCGACATCGCAACAGCTTTCCGCCCGGAACTCGGCTCG 180
Db 17642 GGCGGTTTCGACCGCATCGGACCTCGACATCGGACAGTTCGCGCCCGGAACTCGGCTCG 17701
Qy 181 CGACTGGTTCGTTCAAGCGGGCCCGATGATGCGCCCTACGACGCGCATGAACCGCGCGTTC 240
```


Db	17702	CGACTGGTGTTCACAGCGGGCCGATGATGGCCCTACGACGCCATGAACCGCGCGCTC	17761	TITLE JOURNAL	Direct Submission
Qy	241	GGCGTAGCCACAGGCGAATGGGTACTTTTATAGGCGCGGACGACACCTCTACGAACA	300	REFERENCE AUTHORS	Submitted (16-APR-1999) Microbiology, Colorado State University, Fort Collins, CO 80523-1677, USA 4 (bases 1 to 71286)
Db	17762	GGCGTGGCCACAGGCGAATGGGTACTTTTATAGGCGCGGACGACACCTCTACGAACA	17821	TITLE JOURNAL	Submitted (14-APR-2000) Microbiology, Colorado State University, Fort Collins, CO 80523-1677, USA
Qy	301	ACCAGTGTGGCCAGGTAGCCCTTTTCTCGCGACCATGCGGCAAGCATCTTGTCTAT	360	REMARK COMMENT FEATURES	Sequence update by submitter On Apr 14, 2000 this sequence version replaced gi:5524305. Location/Qualifiers 1. 71286 /organism="Mycobacterium avium" /mol_type="genomic DNA" /strain="2151" /db_xref="taxon:1764" 47. .1132 /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAD44199.1" /db_xref="GI:5524306"
Db	17822	ACCAGTGTGGCCAGGTAGCCCTTTTCTCGCGACCATGCGGCAAGCATCTTGTCTAT	17881	CDS	repeat_region repeat_region CDS
Qy	361	GGCGATGTGTGATCGGTTCGAGAAAGCGGCAATGCGGACCTTTTCACTCGACCGC	420	repeat_region	repeat_region
Db	17882	GGCGATGTGTGATCGGTTCGAGAAAGCGGCAATGCGGACCTTTTCACTCGACCGC	17941	CDS	CDS
Qy	421	CTCCTATTGAGACGCAATTTGTGCCACCAATCGATCTTTACCGCGGTGAGCTTTTCGAC	480	repeat_region	repeat_region
Db	17942	CTCCTATTGAGACGCAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	18001	CDS	CDS
Qy	481	GGCATCGGCGCTTACAACCTGGCTTACCGAGTCTGGCGGACTGGGACTTCAATATTCGC	540	repeat_region	repeat_region
Db	18002	GGCATCGGCGCTTACAACCTGGCTTACCGAGTCTGGCGGACTGGGACTTCAATATTCGC	18061	CDS	CDS
Qy	541	TGCTTCTCAACCCGCGCTGATTACCGCTACATGCGCTGATGATTTCCGAATACAAC	600	repeat_region	repeat_region
Db	18062	TGCTTCTCAACCCGCGCTGATTACCGCTACATGCGCTGATGATTTCCGAATACAAC	18121	CDS	CDS
Qy	601	GACATGACGGGCTTACGATGAGGCGAGGACTGATAAGATTTCAGAAACGGCTGCCA	660	repeat_region	repeat_region
Db	18122	GACATGACGGGCTTACGATGAGGCGAGGACTGATAAGATTTCAGAAACGGCTGCCA	18181	CDS	CDS
Qy	661	ATGTACTTCTGGGTTGCAGGTTGGGAGACTTTCAGCGCATGCTGGCGTTTGTGAAGAC	720	repeat_region	repeat_region
Db	18182	ATGTACTTCTGGGTTGCAGGTTGGGAGACTTTCAGCGCATGCTGGCGTTTGTGAAGAC	18241	CDS	CDS
Qy	721	AAGGAGAAATCGCGCTCGCTACGCGTTGATGAAGGTTTAAGCGCGTCTCCAAA	780	repeat_region	repeat_region
Db	18242	AAGGAGAAATCGCGCTCGCTACGCGTTGATGAAGGTTTAAGCGCGTCTCCAAA	18301	CDS	CDS
Qy	781	GAAACGAGCGCAGACCGTAG	801	repeat_region	repeat_region
Db	18302	GAAACGAGCGCAGACCGTAG	18322	CDS	CDS
RESULT 9	AF143772	71286 bp	DNA	linear	BCT 02-OCT-2003
LOCUS	AF143772				
DEFINITION					
ACCESSION	AF143772				
VERSION	AF143772.2				
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					


```
/protein id="AAD44203.1"
/db_xref="GI:5524310"
/translacion="MTAPHIVDPAGLIGLEALAEASPDLMRSLQITVINALLSADADAV
VGAEGQTPGERAQRNGYRHRGLDRTVGTIDVAPKLRKGTYPDWLLERRKEAES
LITVADCYLAGVSTERNMDKLVKTIGIDSLSKQSVSNATDLDEHVEFRPLDSAG
PFTFVAADALTKVREGGRVNGVILLATGVNGDGRHVRVVTSTGPAWNEFFA
DLVAGLAVGLRVLSDAHAGLEAIAANLPGAORCETHYAANLMSICPKSMPAVK
AMLSVYQDPAPAAQDFRLIDYVTKLPVADLATAERIDEILAFAPKQVMSQI
WSNPAERLKEIRRTDVGIFPNRDAIVRLGVAVLAEQNDWAEGRRLGLDLITR
CRLTTIDNDPDIGANDMPALTA"
repeat_region 5671..5678
/rpt_type=direct
CDS 5791..5876
/notes="similar to Bacteriophage MX1 A-protein encoded by
GenBank Accession Number AF059242"
/codon_start=1
/transl_table=11
/product="A-protein-like protein"
/protein id="AAD44204.1"
/db_xref="GI:5524311"
/translacion="MATELRKSLRRRTSERFGVAMRLARTIKPAETANSSGPLADLA
GPNRPVALFVLGPORSCTSAITVLSICGGLTPTALIGADANNPLGYWEPRRAISLN
EAILRLGTWYDPSFRLDQADESANTACINKIAIYLSITLPAAPLVVKEPRIT
LSDWFOARRAGHDVAAIVIAVRHPQEVICSIAKSVSPALASALMLKCNILAEKNT
RGMERVFYANLLDDMRRETKRISAEPLVELTQDEDAIEKFLTPGLHRQHSQV
RMFGADWTSTVYALRDAHDEPVDVDTATLDRVFDPSYRASERDFRKAHEDFHGLSNML
FRVFRPAVAPVMEIFAMLRNRGTWA"
gene 6996..8057
/genes="gepia"
CDS 6996..8057
/genes="gepia"
/notes="TDP-D-glucose-dehydratase; putative; similar to
Mycobacterium tuberculosis rmlB2 and galE"
/codon_start=1
/transl_table=11
/product="gepia"
/protein id="AAD44205.1"
/db_xref="GI:5524312"
/translacion="MVASVLITGGAGFTGSLSRRLVEAGYDVALMDVLHPQVHGDR
PVELAPSRLFTGVDPHADFDVLRPRPTQIVHLAETGTQASLSRATHGVSNNV
GTTQLLDALSGSLVPEQVLASRNVYEGAGWQSGAEVFPQPRSHQLVAGRWDPQ
GPAGQVAPLSPDRPTPRNTVYATKLAQHLLAAWTAADHNTLSVLQVNYGP
QGSUTNSYTVLVALFARLQOALEVEDGRIVRDFVYIDVVVDALFAVQRPASQ
RLQDVSGRATTIOELANTIAAMCEAPEPVVVGKFRDGDVRAASCIEPATSQLGWH
KWTLEDGLRALLEWIGNRp"
gene 8321..9448
/genes="mtfa"
CDS 8321..9448
/genes="mtfa"
/notes="methyltransferase; putative; similar to mtfb and
mtfC"
/codon_start=1
/transl_table=11
/product="mtfa"
/protein id="AAD44206.1"
/db_xref="GI:5524313"
/translacion="MGIVDRRTVMVPTKLHRAPDRTRHMLWKAVLLGGALLAAACVF
WPVLAVGIGAILLLLCARIQGRDRDVIPLNYARDTFLYDDOYREERTRETLAERL
RIGHTLWESQVLPQGAENSEBELLDLGVWIGNSTRILIFDTRHYVIGDFTSGVL
EDWLEDRIVKRGAPUSLSEPPAQRPIRDTGVINDDGVPAALGRDVRPIKSGTYDLA
PFLADRPAAIRLPHMDJDTYESCLHALETCKDHPVVGSLVDFEYLVINGEMAFYD
FQRYELWQVQYKRAEMENMVTSRKRRLYSIAAIPGYLLGLDGRPLWACFRE
PFRWFLNWAEDIPFLLGAAGSRKSVEITGLKGLAVPH"
gene 9715..10536
/genes="mtfb"
CDS 9715..10536
/genes="mtfb"
/notes="methyltransferase; putative; similar to
methyltransferase mtfC and Micromonospora griseorubida
mycinamin III-O-methyltransferase encoded by GenBank
Accession Number S51595"
/codon_start=1
/transl_table=11
```

```
/product="mtfb"
/protein id="AAD44207.1"
/db_xref="GI:5524314"
/translacion="MTDRMRSPLDTRSAVLLDRLRLTRYSGDELVPVGMVYLGRP
LFSTRKMLVRKPFNAHARDLGDLPADALTMIGQRLTSLOHCVEITLKDIPGDL
VECVWGGASILMRVALSAYDESKRCVMDKLVKTPGPPPTAHYQADQKILHRAAG
ILAVPEAQVRANPERYGLDDRVFPFGWFKDTQADAPIERIAVLRDGDYESTQIA
LDALYPLRLSAGGICIIDYHAIDACROAVTDYRSEHGVTAPIEBIDGTGVLWRKS"
gene 10658..11935
/genes="gtfa"
CDS 10658..11935
/genes="gtfa"
/notes="glycosyltransferase; putative; similar to
rhamnosyltransferase rtfa and glycosyltransferase gtfb"
/codon_start=1
/transl_table=11
/product="gtfa"
/protein id="AAD44208.1"
/db_xref="GI:5524315"
/translacion="MKCVLASYTRGDIPEPVVVVARELQRHGHMIMAVPPDSVSFTE
AAGTADYGLDSOAWLDVYRNFWTSFGRFWMKEMRMWMDLSDOCWQNTT
LLSLADGADILLFAGSQSYOEPAANVAYVDLPLATLHVPMRPNQVSVLPAIARSA
MTAFDWCRLNKKVDAQRELGLPKATGSPRIARERGSLEIQADEVCFPLAAEE
WAKWDGQRPVFGTULMELTTEADDEVMWIRGGTPPICFGFGSPVSPADTVDMIGT
ACAQLGRALLCAGRSDFSGVPAFDHVKVGVPNYATIFPACRAVHHGGSGTAAASL
RAGVPTLILSMDANQTIWGAQLKELKLVGATRRFSATDRDLSLQDLRLKILAPDYARR
```

Query Match	Score	DB 1	Length
Best Local Similarity	99.84	DB 1	71286
Matches	800	Conservative	0
Mismatches	1	Indels	0
Gaps	0		

```
Qy 1 ATGACTGCGGCAGGTGTTCTCGATAATTATCCTACCTTCAATGACGCGGTGACGCTGCAA 60
Db 28153 ATGACTGCGGCAGGTGTTCTCGATAATTATCCTACCTTCAATGACGCGGTGACGCTGCAA 28212

Qy 61 GCCTGCTCGAAGCATCGTGGCGAGACCTTACCGGGAAGTGGAGTGGTCTTGTTCGAC 120
Db 28213 GCCTGCTCGAAGCATCGTGGCGAGACCTTACCGGGAAGTGGAGTGGTCTTGTTCGAC 28272

Qy 121 GGCGGTTTCGACCGATCGGACCTCGACATCGCAACAGTTCCTCGCCCGGAACCTCGGCTCG 180
Db 28273 GGCGGTTTCGACCGATCGGACCTCGACATCGCAACAGTTCCTCGCCCGGAACCTCGGCTCG 28332

Qy 181 CGACTGTTGCTTACAGCGGGCCGATGATGGCCCTTACGACGCGCATGAAACGGGGCGCTC 240
Db 28333 CGACTGTTGCTTACAGCGGGCCGATGATGGCCCTTACGACGCGCATGAAACGGGGCGCTC 28392

Qy 241 GGCGTAGCCACAGCGGAATGGGTACTTTTTTATGGCGCGGACGACACCTCTACGAACCA 300
Db 28393 GGCGTAGCCACAGCGGAATGGGTACTTTTTTATGGCGCGGACGACACCTCTACGAACCA 28452

Qy 301 ACCAGTTGGCCAGGTAGCGCTTTTCTGGCGACCATGCGGAAGCCATCTTGTCTAT 360
Db 28453 ACCAGTTGGCCAGGTAGCGCTTTTCTGGCGACCATGCGGAAGCCATCTTGTCTAT 28512

Qy 361 GGCGATGTTGTGATGCGTTCGACGAAAAGCCGCGCATGCGGACCTTTCGACCTCGACGCG 420
Db 28513 GGCGATGTTGTGATGCGTTCGACGAAAAGCCGCGCATGCGGACCTTTCGACCTCGACGCG 28572

Qy 421 CTCCTATTGAGACGAATTTGTGCGCAATTCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
Db 28573 CTCCTATTGAGACGAATTTGTGCGCAATTCGATCTTTTACCGCGGTGAGCTTTTCGAC 28632

Qy 481 GGCAATCGGCTTACACCTCGCTACCGAGTCTGGCGGACTGGGACTTCAATATTTCGC 540
Db 28633 GGCAATCGGCTTACACCTCGCTACCGAGTCTGGCGGACTGGGACTTCAATATTTCGC 28692

Qy 541 TGCTTCTCCAAACCGGCGGTGATTACCGCTTACATGACGCTCGTGAATTCGGAATACAAAC 600
Db 28693 TGCTTCTCCAAACCGGCGGTGATTACCGCTTACATGACGCTCGTGAATTCGGAATACAAAC 28752

Qy 601 GACATGACCGGCTTACGATGAGCAGGGGACTGATAAAGAGTTTCAGAAAACGGCTGCA 660
Db 601 GACATGACCGGCTTACGATGAGCAGGGGACTGATAAAGAGTTTCAGAAAACGGCTGCA 660
```

```
Db 28753 GACATGACCGGCTTCAGCATGAGGAGGAGCTGATGAAGAGTTTCAGAAAACGGCTGCCA 28812
Qy 661 ATGTAATCTTGGTTGCGAGGTGGAGAGCTTCAGGCGCATGCTGGCGTTTGTGAAGAC 720
Db 28813 ATGTAATCTTGGTTGCGAGGTGGAGAGCTTCAGGCGCATGCTGGCGTTTGTGAAGAC 28872
Qy 721 AAGGAGAAATCGCGTCTGGCTTGGTACGGGTTGATAAGGGTTAAAGGCGCTCTCCAAA 780
Db 28873 AAGGAGAAATCGCGTCTGGCTTGGTACGGGTTGATAAGGGTTAAAGGCGCTCTCCAAA 28932
Qy 781 GAACGAAGCGAGAACCGTAG 801
Db 28933 GAACGAAGCGAGAACCGTAG 28953

RESULT 10
LOCUS A63817 828 bp DNA linear PAT 12-MAR-1998
DEFINITION Sequence 38 from Patent WO9723624.
ACCESSION A63817
VERSION A63817.1 GI:3717387
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
REFERENCE
AUTHORS Hermon-Taylor,J., Doran,T., Millar,D., Tizard,M., Loughlin,M.,
Sumar,N. and Ford,J.
TITLE NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA
AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY
JOURNAL Patent: WO 9723624-A 38 03-JUL-1997;
ST GEORGE S HOSPITAL MEDICAL S (GB)
COMMENT Other publication AU 120797 19970717.
FEATURES
source Location/Qualifiers
1..828
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
CDS
1..828
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA03755.1"
/db_xref="GI:3717388"
/translation="MVQPKRYAGLTAAATKKVMAAPMFSIIIPLTNVAAVLPACLDLS
IARQTCGFELVLDGSDTDELDIANIAPENLGERLIHHRDQGVDAWNRGVDLA
TGTWLLFLGADDSLYEADTLARVAAFIHEPESDLVYGDVIMRSTNFRWGGAPDLRL
LFKRNICHOAIFYRGLFTIGPNLRYRLADWDFNIRCFSPALVTRYMHVVVASY
NEFGGLSNTIIVDKELKRLPMSTRGLRLVILVRRWPKVISRAMVWRTVLSWRRR"

Query Match 43.2%; Score 346; DB 6; Length 828;
Best Local Similarity 71.0%; Pred. No. 3.2e-63;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;
Qy 1 ATGACTCGCGCAGTGTCTCGATAATTATCCCTACCTTTCAATGACGAGCGTGACGCTGCAA 60
Db 58 ATGGCGCGCACCAATGTTTCGATCATCATCCCCACCTTGACGCTGCGGTATGCT 117
Qy 61 GCCTGCCTCGGAAGCATCGTCGGGCGAGACCTACCGGGGAAGTGGAAAGTGGTCTCTGTCGAC 120
Db 118 GCCTGCCTCGACAGCATCGCCGTCGAGACCTGCGGTGACTTCGAGCTGCTACTGCTGCGAC 177
Qy 121 GCGGTTTCGACCGATCGGACCTCGACATCGGGAACAGTTCGCGCCCGGAACCTCGGCTCG 180
Db 178 GCGGCTCGACGAGGAGAAACCTCGACATCGCAACATTTTGGCCCCCAACCTCGGCGAG 237
Qy 181 GCACTGTTGTTTCACAGGGGGCCCGATGATGSCCCTACGACGCCATGAACCGGGCGCTC 240
Db 238 CGGTTGATCATTCATCGGACACCGACGAGGCGTCTACGACGCCATGAACCGGCGGTG 297
Qy 241 GCGGTAGCCACAGGGGAATGGGTACTTTTTTTAGCGCGCGAGCAGACACCTCTACGAACCA 300
Db 298 GACCTGGCACCGGAACGTTGGTCTCTTCTGCGGCGGCGGACGACAGCTGTACGAGGCT 357
```

```
Qy 301 ACCAGTTGGCCAGGTAGCGCTTTTCTCGGCCACCATGCGGAAGCCATCTTGTCTCAT 360
Db 358 GACACCTTGGCGGGTGGCGCCTTCATTGGCGAAACAGAGCCAGCGATCTGATATAT 417
Qy 361 GGCATGTTGTGATGTCGCTTCGACGAAAAGCCGGCATGCGGACCTTTTCGACCTCGACGCG 420
Db 418 GCGACGTGATCATGCGCTCAACCAATTCGCTGGGGTGGCGCTTCGACCTCGACCGT 477
Qy 421 CTCCTATTGTGAGACGAATTTGTGCCCAATTCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
Db 478 CTGTTGTTCAAGCGCAACATCTGCCATCAGCGCATCTTACCGCGCGGACTCTTCGCGC 537
Qy 481 GGCATCGGCCCTTACAACTCGCTACCGAGTCTGGGGGAGCTGGGACTTCAATATTTCGC 540
Db 538 ACCATCGGTCCCTACAACCTCCGCTACCGGTCCTCGCGAGCTTCAATATTTCGC 597
Qy 541 TGCCTTCTCAACCCGGCGCTGATTACCGCTACATGACGCTCGTGATTTCCGAATACAAC 600
Db 598 TGCCTTCTCAACCCAGCGCTCGTCACCGCTACATGCACTGCTGCTTGGCAAGCTACAAC 657
Qy 601 GACATGACCGGCTTCAGCATGAGGAGGAGCTGATAAAGAGTTTCAGAAAACGGCTGCCA 660
Db 658 GAATTGCGCGGCTCAGCAATACGATCG---TCGACAAAGAGTTTTCGAGCGGCTGCCG 714
Qy 661 ATGTAC 666
Db 715 ATGTCC 720

RESULT 11
LOCUS MTU00024/c 31175 bp DNA linear BCT 11-MAR-2002
DEFINITION Mycobacterium tuberculosis cosmid tbc2.
ACCESSION U00024
VERSION U00024.1 GI:560506
KEYWORDS
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE
AUTHORS Smith,D.R.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 31175)
AUTHORS Robison,K.
TITLE Direct Submission
JOURNAL Submitted (29-SEP-1994) Department of Genetics, Harvard Medical
School, 200 Longwood Avenue, Boston MA 02115
COMMENT On Oct 27, 1994 this sequence version replaced gi:414230.
This sequence data was produced by the Genome Sequencing Center
located at Genome Therapeutics Inc. (formerly Collaborative
Research Inc.) (100 Beaver St., Waltham MA, 02154 617-893-5007).
Please contact Doug Smith (smith@cric.com) for further
information. The annotation should be considered preliminary and
incomplete.
FEATURES
source Location/Qualifiers
1..31175
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"
<1..902
/codon_start=3
/transl_table=11
/product="u0002a"
/db_xref="GI:560529"
/transl_table="SPRCARTLAGFESRLACRRYARGAVALTAATSQRIEVRISSET
VADSKPPRDLKSDGSFRLVQIGTTIVVFAVVFYVITSRDDKDGVAAGPG
DAVRTSKLVTPQTSNPKAVSFYDFELCPACGIFERGPGTIVSKLVDIGAADY
TMVAILDASNQHYSSRAAAAYCADESIEAFRRFHAALFSDIQPAELGKDFDNA
RLIELAREAGVVGKVPDCINSKYIEKVDGLAAAVNVHATPTVRVNGTEYEWSTPAAL
```

```
CDS
VAKIKIEIVGDVPGIDSAATATS"
908..1540
/codon_start=1
/transl_table=11
/product="u0002b"
/protein_id="AAAS0949.1"
/db_xref="GI:560528"
/translacion="MVAARPAERSGDPAARVPVPSAWVVLIGVGIGLFPASMTLTVK
VRLDPIYVSCNVPYVSCSVNTTQASLLGPNPLLAGTFTVVVTVGLAVAK
VLLPWWYGLAVGLVGVFVHLLIQLSRLVIGALCPYCMVMAVATLTVVASIV
FGPMRENGSERVGARLLYQWRWSLATLMTFTVLLIMVRFWDYWSLTI"
1565..4939
/codon_start=1
/transl_table=11
/product="pvc"
/protein_id="AAAS0948.1"
/db_xref="GI:560527"
/translacion="NFSKVLVANRGEIAIRAFRAAYELGVGTAVYVYEDRNSQHLK
ADESQIGDIGHVHAYLSVDEIVATARRAGADAIYPGYFELSENPDIAAACAAGIS
FVGPSEVLELAGNKSRAIAAREAGLFLMSAFPSVDELLSVAAAGMPFLFYKAV
AGGGGRMRRVGDIAALPEATEAASREAESAFGDPVTYLEQAVINPHIYQIILADNL
GDVHLVPRDCSVORRHQVLELAPAPHLDAELRYKMCVDAVAFARHIGYSCAGTVEF
LLDEGEYVFIEMPRVQVHTVTEITDVLVASQLAIAAGETLEQLGLQEBIAPH
GAAQCRIITEDPANGFRPRAGSARCDDPAVPVSAWTAAPTWRKNQVRLRHAGQAD
LSGRDLTPASRRARIAEPRIRGVSTNIPFLQAVLDDPDPFRAGVTTSFIDERPOL
TARASADRGTKILNPLADVTNNPYGSRPSTIYDDKLPDLDLRAAPAGSKQRLVKL
GPEGFARLRESAAVGVDTTFRDAHQSLATRVTSGLSVAPYLAKMTQQLLSVEC
WGGATYDVALRFLKEDPWERLATRAAMNPLICMMLRGNTVGTVPYPIVTSAFVQ
EATATGIDIFRIDALNIESRPAIDAVRETSNAIEAVMCMYCGDLTDGCEQLYTL
TYLKLAEQIVDAGHVLAIKDMAGLLRPPAQRVLSURSPFDLPVHLHTHDPGQL
ASYAAMHAGADAVGAAPLAGTTSQPALSSI VAAAHTEYDGLSLVCALEPYW
EALARKVAFESGLPGFTGRVYHHEIPGQSLNLRQQAIALGLGDRFEEIIEAGVAG
RVGLRLVPTKVVGDALALVAGVSDAFASDPARFGIPESVLFGELRGELGDDP
G3WPEPLTAALGAARPTAOLAAADDEIALSSVGAKQAKTLNELLPPSPKSENEH
REAGDTSQSANQFPYGLRGEEHVRKLERGVELLIGLEAISPEDEGRMTVMCILN
QOLRPVLTRPISASAVPAEAKDRNGEHTAAPAGVTVGVVCVGERVAGAGQTIATI
EAMKMEAPITPAVAGTVERVAVSDTAEWRAETCWWW"
5057..5704
/codon_start=1
/transl_table=11
/product="u0002d"
/protein_id="AAAS0947.1"
/db_xref="GI:560526"
/translacion="MQGYRPLAGDEAMRWGRYPLARESGLDHPHRRRRRIAVP
PRGRTPTDTRVNFIVTARRDLTGLAVLDLVAGSGALGLEALSRAAGASVLFVESD
QRSAAVIARNEALGLSGATLRRCGAVAAVAAGTTSVDVLADPPYVNDVADVDAIL
AALGTNGTREGTVAVVERATTCAPLTWPEGWRWRPQRYVYGDTRLELAERLFANV"
5790..6275
/codon_start=1
/transl_table=11
/product="u0002e"
/protein_id="AAAS0946.1"
/db_xref="GI:560525"
/translacion="MTGAVCPGSDPDPVTLGHVDIFERAAAQFDEVVAAILVNPAGTM
QMKDRIAMVKESTHPLNLRVQVGHGLVVDVFRSCGMTAIVKGLTGCTDPEYELQWA
FQNLHIAQVDVTFVATAPRYSFVSSSLAKEVAMLGGDVSELLPEPVNRRRLDRLNT
T"
complement(7144..8076)
/codon_start=1
/transl_table=11
/product="purU"
/protein_id="AAAS0945.1"
/db_xref="GI:560524"
/translacion="WKGKSWTAHATPNEDYPDPGPPGPPADIGRLLLRCHDRPGII
AAVSTFLARAGANIISLDQHPATEGGTFLORAIFFPLGLTAAVDELDQRDGGSTVADK
FGIDYRFAAARAKRVAIMASTEDHCLDLRLNRNGELEMVVMVIANHPDLAAHVR
PFGVPFTHIPATRTDTEAQRQLLSGNVDLVLYARMQILSPGLEAIGCPLINI
HHSFLPAFTGAAPQORARERGVKLIGATAYVTVLDEGPIIEQDVVRVDHTHTVDLL
VRVGADVERAVLSRAVLWHCCODRVIVHNNQIVF"
complement(8149..9369)
/codon_start=1
/transl_table=11
/product="u0002f"
/protein_id="AAAS0944.1"
/db_xref="GI:560523"
/translacion="WTSTKVEDRVTAAVLGAIGHALALTASMTWILWALILGFALSA
VVOAVRRESTITVLLGDDRPRTLATATGLGAASSCSVAVALASLFRKGANFTAAAM
APEIGSTNLVVELGIILALLMGWQFTAAEFVGGPMILVLAFLRFLFGARLIDAARE
QAERGLAGSMEGHAAMDSIKREGSFWRRLSLPGFTSIASHVFMELAILRDLILGL
LIRAGIAAWPESFQWQFFFLANHPAWSAVMGPIIGPIVAIVSFVCSIGNVPLAAVLWN
GIGIFGGVIAFIFADLLIPLINIRYKTYGARMMLVLGLTFYASVMVAGYITELLEFGT
TNLIPORSASTMTAEISWNTYTLNVLFLVIAAALVVRFTTSGGLPMLRMWGGSPDA
PHLHDDRHDDHLGH"
9306..10433
/codon_start=1
/transl_table=11
/product="u0002ja"
/protein_id="AAAS0943.1"
/db_xref="GI:560522"
/translacion="WRASRLPALLPSLGPPLWLSRSALHHIPGCGIYTVGNDRAVV
PLNVSPHSCGFPFRVRSVYATASRSGGPPVASEVRGDAISTTPDAAPGLAARRR
RILFVAVTLAHVVRFPALAQSLDPSRYEHPACDPYNNQLLGLPLPRHHAHTIPS
ERFNGNLTQGRFYAMRTLRKYVEADRLVDEIAPDLVVGDLRISLSVSRALAGIPIA
IANAYWSPYAORRFPPLDVIWTRLFGVRLVKLLYRLRPLLPALQCMPLNWMVRRHGL
SSLGNLCRIETDGDHTLYADVPELMTYDLPANHEVILGPVLMSPAGPPTPWDSLPT
DRPIVATYLTGSGGRNLLQLVLNLAELPVTVIAATAAATV"
10439..10846
/codon_start=1
/transl_table=11
/product="u0002h"
/protein_id="AAAS0942.1"
/db_xref="GI:560521"
/translacion="MPANAFVADYLPGEAAAARSAVVVNCNGSLTTOQALVAGVPVIG
VAGNLDQHLNMEAVRACAGVLLRTERLKSQRVAGAVMQVISRSEYPTSRATRCLR
TRPWIIPAAECGCSAPHAGEPAPDMVGQLVTS"
11766..12503
/codon_start=1
/transl_table=11
/product="u0002jb"
/protein_id="AAAS0941.1"
/db_xref="GI:560520"
/translacion="MGLVWRSRTSLVGLIGLVRVASFAAQLFYRPSDAVAEEVHKW
YGNLWTKTYMGINCWKSVDMMNYQEILSELOPSLVIEFGTRYGGSAYVAFANMR
QIGQPFKVLTVDNSHKALDPRARREPDLFVSESSSTDPAIAEQRLKNEYFGKIFAI
LDSHSMNVHLAEMKLLRPLLSAGDYLVDVESDINGHPVLPFGFGPGYEAEIAYBDEF
PNDYKDAERENKFGWTSAPNGFLIRN"
12604..13995
/codon_start=1
/transl_table=11
/product="u0002ka"
/protein_id="AAAS0940.1"
/db_xref="GI:560519"
/translacion="MEETSAGDPDPDAGTSTAPNAPEPVARQRILFVGEAATLAH
VREFVLARSIDPSRYEHPACDPRENKLLGLPLPPHPIHTVPSEVLLKIAOGRLF
YNTFLRKTYAADRKIINEIADPVVVDGNRLSLVSARLAGIPYIAIANATWSPARR
RFPDPDVPTRFFRGVRPVSIILYRLRYLFALYCLPLNWLRRKHLSLGLWDLCKRIFT
DGDYTLADVPELVTYNTLPANHYLGPVLMSPDPKPTWWHSLPTDRPIIYATIGSS
GSKNLLQVNALADLPVTIAATAGRNHLKNVPANAFVADLPGEAAAARLAVVLCN
GGSPTQOALAAAGVPVIGLPSMDOHLNMEALERAGAVLLRTERLNTGTEGVAAAQVQ
LSGAEFRQAAERLPPKPSDQTLPSDRTSKVRCGSSAKGCPYRRLANGCPVEHLIRER
VAVNAEVPQRLS"
complement(14306..15133)
/codon_start=1
/transl_table=11
/product="u0002kc"
/protein_id="AAAS0938.1"
/db_xref="GI:560517"
/translacion="MVQTKRYAGLTANTKKVAAWAPMFSIIIPTLNVAAYLPACLDS
IARCTCGDFELVLDGGSDBETLDIANIFAPNLGERLIIRHDTQGVVDANRNGVDLA
TKENILFLGADDSLYEARGLTARVAATFGEHPSDLVYGDVIMRSTNFRMGCAFOLDRL
LFKNNIQAIFVYRGLFGTIGPVNARVLAVDWDFNTRCFSPNALVRYHVVVVASV
NEFGGLSNTIVDEKFLKRLPMSTFLGRLVILVVRPKVISRAMWRTVISWRRR"
complement(15203..15934)
/codon_start=1
/transl_table=11
/product="u0002kb"
/protein_id="AAAS0944.1"
/db_xref="GI:560523"
/translacion="WTSTKVEDRVTAAVLGAIGHALALTASMTWILWALILGFALSA
VVOAVRRESTITVLLGDDRPRTLATATGLGAASSCSVAVALASLFRKGANFTAAAM
APEIGSTNLVVELGIILALLMGWQFTAAEFVGGPMILVLAFLRFLFGARLIDAARE
QAERGLAGSMEGHAAMDSIKREGSFWRRLSLPGFTSIASHVFMELAILRDLILGL
LIRAGIAAWPESFQWQFFFLANHPAWSAVMGPIIGPIVAIVSFVCSIGNVPLAAVLWN
GIGIFGGVIAFIFADLLIPLINIRYKTYGARMMLVLGLTFYASVMVAGYITELLEFGT
TNLIPORSASTMTAEISWNTYTLNVLFLVIAAALVVRFTTSGGLPMLRMWGGSPDA
PHLHDDRHDDHLGH"
9306..10433
/codon_start=1
/transl_table=11
/product="u0002ja"
/protein_id="AAAS0943.1"
/db_xref="GI:560522"
/translacion="WRASRLPALLPSLGPPLWLSRSALHHIPGCGIYTVGNDRAVV
PLNVSPHSCGFPFRVRSVYATASRSGGPPVASEVRGDAISTTPDAAPGLAARRR
RILFVAVTLAHVVRFPALAQSLDPSRYEHPACDPYNNQLLGLPLPRHHAHTIPS
ERFNGNLTQGRFYAMRTLRKYVEADRLVDEIAPDLVVGDLRISLSVSRALAGIPIA
IANAYWSPYAORRFPPLDVIWTRLFGVRLVKLLYRLRPLLPALQCMPLNWMVRRHGL
SSLGNLCRIETDGDHTLYADVPELMTYDLPANHEVILGPVLMSPAGPPTPWDSLPT
DRPIVATYLTGSGGRNLLQLVLNLAELPVTVIAATAAATV"
10439..10846
/codon_start=1
/transl_table=11
/product="u0002h"
/protein_id="AAAS0942.1"
/db_xref="GI:560521"
/translacion="MPANAFVADYLPGEAAAARSAVVVNCNGSLTTOQALVAGVPVIG
VAGNLDQHLNMEAVRACAGVLLRTERLKSQRVAGAVMQVISRSEYPTSRATRCLR
TRPWIIPAAECGCSAPHAGEPAPDMVGQLVTS"
11766..12503
/codon_start=1
/transl_table=11
/product="u0002jb"
/protein_id="AAAS0941.1"
/db_xref="GI:560520"
/translacion="MGLVWRSRTSLVGLIGLVRVASFAAQLFYRPSDAVAEEVHKW
YGNLWTKTYMGINCWKSVDMMNYQEILSELOPSLVIEFGTRYGGSAYVAFANMR
QIGQPFKVLTVDNSHKALDPRARREPDLFVSESSSTDPAIAEQRLKNEYFGKIFAI
LDSHSMNVHLAEMKLLRPLLSAGDYLVDVESDINGHPVLPFGFGPGYEAEIAYBDEF
PNDYKDAERENKFGWTSAPNGFLIRN"
12604..13995
/codon_start=1
/transl_table=11
/product="u0002ka"
/protein_id="AAAS0940.1"
/db_xref="GI:560519"
/translacion="MEETSAGDPDPDAGTSTAPNAPEPVARQRILFVGEAATLAH
VREFVLARSIDPSRYEHPACDPRENKLLGLPLPPHPIHTVPSEVLLKIAOGRLF
YNTFLRKTYAADRKIINEIADPVVVDGNRLSLVSARLAGIPYIAIANATWSPARR
RFPDPDVPTRFFRGVRPVSIILYRLRYLFALYCLPLNWLRRKHLSLGLWDLCKRIFT
DGDYTLADVPELVTYNTLPANHYLGPVLMSPDPKPTWWHSLPTDRPIIYATIGSS
GSKNLLQVNALADLPVTIAATAGRNHLKNVPANAFVADLPGEAAAARLAVVLCN
GGSPTQOALAAAGVPVIGLPSMDOHLNMEALERAGAVLLRTERLNTGTEGVAAAQVQ
LSGAEFRQAAERLPPKPSDQTLPSDRTSKVRCGSSAKGCPYRRLANGCPVEHLIRER
VAVNAEVPQRLS"
complement(14306..15133)
/codon_start=1
/transl_table=11
/product="u0002kc"
/protein_id="AAAS0938.1"
/db_xref="GI:560517"
/translacion="MVQTKRYAGLTANTKKVAAWAPMFSIIIPTLNVAAYLPACLDS
IARCTCGDFELVLDGGSDBETLDIANIFAPNLGERLIIRHDTQGVVDANRNGVDLA
TKENILFLGADDSLYEARGLTARVAATFGEHPSDLVYGDVIMRSTNFRMGCAFOLDRL
LFKNNIQAIFVYRGLFGTIGPVNARVLAVDWDFNTRCFSPNALVRYHVVVVASV
NEFGGLSNTIVDEKFLKRLPMSTFLGRLVILVVRPKVISRAMWRTVISWRRR"
complement(15203..15934)
/codon_start=1
/transl_table=11
/product="u0002kb"
/protein_id="AAAS0944.1"
/db_xref="GI:560523"
/translacion="WTSTKVEDRVTAAVLGAIGHALALTASMTWILWALILGFALSA
VVOAVRRESTITVLLGDDRPRTLATATGLGAASSCSVAVALASLFRKGANFTAAAM
APEIGSTNLVVELGIILALLMGWQFTAAEFVGGPMILVLAFLRFLFGARLIDAARE
QAERGLAGSMEGHAAMDSIKREGSFWRRLSLPGFTSIASHVFMELAILRDLILGL
LIRAGIAAWPESFQWQFFFLANHPAWSAVMGPIIGPIVAIVSFVCSIGNVPLAAVLWN
GIGIFGGVIAFIFADLLIPLINIRYKTYGARMMLVLGLTFYASVMVAGYITELLEFGT
TNLIPORSASTMTAEISWNTYTLNVLFLVIAAALVVRFTTSGGLPMLRMWGGSPDA
PHLHDDRHDDHLGH"
9306..10433
/codon_start=1
/transl_table=11
/product="u0002ja"
/protein_id="AAAS0943.1"
/db_xref="GI:560522"
/translacion="WRASRLPALLPSLGPPLWLSRSALHHIPGCGIYTVGNDRAVV
PLNVSPHSCGFPFRVRSVYATASRSGGPPVASEVRGDAISTTPDAAPGLAARRR
RILFVAVTLAHVVRFPALAQSLDPSRYEHPACDPYNNQLLGLPLPRHHAHTIPS
ERFNGNLTQGRFYAMRTLRKYVEADRLVDEIAPDLVVGDLRISLSVSRALAGIPIA
IANAYWSPYAORRFPPLDVIWTRLFGVRLVKLLYRLRPLLPALQCMPLNWMVRRHGL
SSLGNLCRIETDGDHTLYADVPELMTYDLPANHEVILGPVLMSPAGPPTPWDSLPT
DRPIVATYLTGSGGRNLLQLVLNLAELPVTVIAATAAATV"
10439..10846
/codon_start=1
/transl_table=11
/product="u0002h"
/protein_id="AAAS0942.1"
/db_xref="GI:560521"
/translacion="MPANAFVADYLPGEAAAARSAVVVNCNGSLTTOQALVAGVPVIG
VAGNLDQHLNMEAVRACAGVLLRTERLKSQRVAGAVMQVISRSEYPTSRATRCLR
TRPWIIPAAECGCSAPHAGEPAPDMVGQLVTS"
11766..12503
/codon_start=1
/transl_table=11
/product="u0002jb"
/protein_id="AAAS0941.1"
/db_xref="GI:560520"
/translacion="MGLVWRSRTSLVGLIGLVRVASFAAQLFYRPSDAVAEEVHKW
YGNLWTKTYMGINCWKSVDMMNYQEILSELOPSLVIEFGTRYGGSAYVAFANMR
QIGQPFKVLTVDNSHKALDPRARREPDLFVSESSSTDPAIAEQRLKNEYFGKIFAI
LDSHSMNVHLAEMKLLRPLLSAGDYLVDVESDINGHPVLPFGFGPGYEAEIAYBDEF
PNDYKDAERENKFGWTSAPNGFLIRN"
12604..13995
/codon_start=1
/transl_table=11
/product="u0002ka"
/protein_id="AAAS0940.1"
/db_xref="GI:560519"
/translacion="MEETSAGDPDPDAGTSTAPNAPEPVARQRILFVGEAATLAH
VREFVLARSIDPSRYEHPACDPRENKLLGLPLPPHPIHTVPSEVLLKIAOGRLF
YNTFLRKTYAADRKIINEIADPVVVDGNRLSLVSARLAGIPYIAIANATWSPARR
RFPDPDVPTRFFRGVRPVSIILYRLRYLFALYCLPLNWLRRKHLSLGLWDLCKRIFT
DGDYTLADVPELVTYNTLPANHYLGPVLMSPDPKPTWWHSLPTDRPIIYATIGSS
GSKNLLQVNALADLPVTIAATAGRNHLKNVPANAFVADLPGEAAAARLAVVLCN
GGSPTQOALAAAGVPVIGLPSMDOHLNMEALERAGAVLLRTERLNTGTEGVAAAQVQ
LSGAEFRQAAERLPPKPSDQTLPSDRTSKVRCGSSAKGCPYRRLANGCPVEHLIRER
VAVNAEVPQRLS"
complement(14306..15133)
/codon_start=1
/transl_table=11
/product="u0002kc"
/protein_id="AAAS0938.1"
/db_xref="GI:560517"
/translacion="MVQTKRYAGLTANTKKVAAWAPMFSIIIPTLNVAAYLPACLDS
IARCTCGDFELVLDGGSDBETLDIANIFAPNLGERLIIRHDTQGVVDANRNGVDLA
TKENILFLGADDSLYEARGLTARVAATFGEHPSDLVYGDVIMRSTNFRMGCAFOLDRL
LFKNNIQAIFVYRGLFGTIGPVNARVLAVDWDFNTRCFSPNALVRYHVVVVASV
NEFGGLSNTIVDEKFLKRLPMSTFLGRLVILVVRPKVISRAMWRTVISWRRR"
complement(15203..15934)
/codon_start=1
/transl_table=11
/product="u0002kb"
/protein_id="AAAS0944.1"
/db_xref="GI:560523"
/translacion="WTSTKVEDRVTAAVLGAIGHALALTASMTWILWALILGFALSA
VVOAVRRESTITVLLGDDRPRTLATATGLGAASSCSVAVALASLFRKGANFTAAAM
APEIGSTNLVVELGIILALLMGWQFTAAEFVGGPMILVLAFLRFLFGARLIDAARE
QAERGLAGSMEGHAAMDSIKREGSFWRRLSLPGFTSIASHVFMELAILRDLILGL
LIRAGIAAWPESFQWQFFFLANHPAWSAVMGPIIGPIVAIVSFVCSIGNVPLAAVLWN
GIGIFGGVIAFIFADLLIPLINIRYKTYGARMMLVLGLTFYASVMVAGYITELLEFGT
TNLIPORSASTMTAEISWNTYTLNVLFLVIAAALVVRFTTSGGLPMLRMWGGSPDA
PHLHDDRHDDHLGH"
9306..10433
/codon_start=1
/transl_table=11
/product="u0002ja"
/protein_id="AAAS0943.1"
/db_xref="GI:560522"
/translacion="WRASRLPALLPSLGPPLWLSRSALHHIPGCGIYTVGNDRAVV
PLNVSPHSCGFPFRVRSVYATASRSGGPPVASEVRGDAISTTPDAAPGLAARRR
RILFVAVTLAHVVRFPALAQSLDPSRYEHPACDPYNNQLLGLPLPRHHAHTIPS
ERFNGNLTQGRFYAMRTLRKYVEADRLVDEIAPDLVVGDLRISLSVSRALAGIPIA
IANAYWSPYAORRFPPLDVIWTRLFGVRLVKLLYRLRPLLPALQCMPLNWMVRRHGL
SSLGNLCRIETDGDHTLYADVPELMTYDLPANHEVILGPVLMSPAGPPTPWDSLPT
DRPIVATYLTGSGGRNLLQLVLNLAELPVTVIAATAAATV"
10439..10846
/codon_start=1
/transl_table=11
/product="u0002h"
/protein_id="AAAS0942.1"
/db_xref="GI:560521"
/translacion="MPANAFVADYLPGEAAAARSAVVVNCNGSLTTOQALVAGVPVIG
VAGNLDQHLNMEAVRACAGVLLRTERLKSQRVAGAVMQVISRSEYPTSRATRCLR
TRPWIIPAAECGCSAPHAGEPAPDMVGQLVTS"
11766..12503
/codon_start=1
/transl_table=11
/product="u0002jb"
/protein_id="AAAS0941.1"
/db_xref="GI:560520"
/translacion="MGLVWRSRTSLVGLIGLVRVASFAAQLFYRPSDAVAEEVHKW
YGNLWTKTYMGINCWKSVDMMNYQEILSELOPSLVIEFGTRYGGSAYVAFANMR
QIGQPFKVLTVDNSHKALDPRARREPDLFVSESSSTDPAIAEQRLKNEYFGKIFAI
LDSHSMNVHLAEMKLLRPLLSAGDYLVDVESDINGHPVLPFGFGPGYEAEIAYBDEF
PNDYKDAERENKFGWTSAPNGFLIRN"
12604..13995
/codon_start=1
/transl_table=11
/product="u0002ka"
/protein_id="AAAS0940.1"
/db_xref="GI:560519"
/translacion="MEETSAGDPDPDAGTSTAPNAPEPVARQRILFVGEAATLAH
VREFVLARSIDPSRYEHPACDPRENKLLGLPLPPHPIHTVPSEVLLKIAOGRLF
YNTFLRKTYAADRKIINEIADPVVVDGNRLSLVSARLAGIPYIAIANATWSPARR
RFPDPDVPTRFFRGVRPVSIILYRLRYLFALYCLPLNWLRRKHLSLGLWDLCKRIFT
DGDYTLADVPELVTYNTLPANHYLGPVLMSPDPKPTWWHSLPTDRPIIYATIGSS
GSKNLLQVNALADLPVTIAATAGRNHLKNVPANAFVADLPGEAAAARLAVVLCN
GGSPTQOALAAAGVPVIGLPSMDOHLNMEALERAGAVLLRTERLNTGTEGVAAAQVQ
LSGAEFRQAAERLPPKPSDQTLPSDRTSKVRCGSSAKGCPYRRLANGCPVEHLIRER
VAVNAEVPQRLS"
complement(14306..15133)
/codon_start=1
/transl_table=11
/product="u0002kc"
/protein_id="AAAS0938.1"
/db_xref="GI:560517"
/translacion="MVQTKRYAGLTANTKKVAAWAPMFSIIIPTLNVAAYLPACLDS
IARCTCGDFELVLDGGSDBETLDIANIFAPNLGERLIIRHDTQGVVDANRNGVDLA
TKENILFLGADDSLYEARGLTARVAATFGEHPSDLVYGDVIMRSTNFRMGCAFOLDRL
LFKNNIQAIFVYRGLFGTIGPVNARVLAVDWDFNTRCFSPNALVRYHVVVVASV
NEFGGLSNTIVDEKFLKRLPMSTFLGRLVILVVRPKVISRAMWRTVISWRRR"
complement(15203..15934)
/codon_start=1
/transl_table=11
/product="u0002kb"
/protein_id="AAAS0944.1"
/db_xref="GI:560523"
/translacion="WTSTKVEDRVTAAVLGAIGHALALTASMTWILWALILGFALSA
VVOAVRRESTITVLLGDDRPRTLATATGLGAASSCSVAVALASLFRKGANFTAAAM
APEIGSTNLVVELGIILALLMGWQFTAAEFVGGPMILVLAFLRFLFGARLIDAARE
QAERGLAGSMEGHAAMDSIKREGSFWRRLSLPGFTSIASHVFMELAILRDLILGL
LIRAGIAAWPESFQWQFFFLANHPAWSAVMGPIIGPIVAIVSFVCSIGNVPLAAVLWN
GIGIFGGVIAFIFADLLIPLINIRYKTYGARMMLVLGLTFYASVMVAGYITELLEFGT
TNLIPORSASTMTAEISWNTYTLNVLFLVIAAALVVRFTTSGGLPMLRMWGGSPDA
PHLHDDRHDDHLGH"
9306..10433
/codon_start=1
/transl_table=11
/product="u0002ja"
/protein_id="AAAS0943.1"
/db_xref="GI:560522"
/translacion="WRASRLPALLPSLGPPLWLSRSALHHIPGCGIYTVGNDRAVV
PLNVSPHSCGFPFRVRSVYATASRSGGPPVASEVRGDAISTTPDAAPGLAARRR
RILFVAVTLAHVVRFPALAQSLDPSRYEHPACDPYNNQLLGLPLPRHHAHTIPS
ERFNGNLTQGRFYAMRTLRKYVEADRLVDEIAPDLVVGDLRISLSVSRALAGIPIA
IANAYWSPYAORRFPPLDVIWTRLFGVRLVKLLYRLRPLLPALQCMPLNWMVRRHGL
SSLGNLCRIETDGDHTLYADVPELMTYDLPANHEVILGPVLMSPAGPPTPWDSLPT
DRPIVATYLTGSGGRNLLQLVLNLAELPVTVIAATAAATV"
10439..10846
/codon_start=1
/transl_table=11
/product="u0002h"
/protein_id="AAAS0942.1"
/db_xref="GI:560521"
/translacion="MPANAFVADYLPGEAAAARSAVVVNCNGSLTTOQALVAGVPVIG
VAGNLDQHLNMEAVRACAGVLLRTERLKSQRVAGAVMQVISRSEYPTSRATRCLR
TRPWIIPAAECGCSAPHAGEPAPDMVGQLVTS"
11766..12503
/codon_start=1
/transl_table=11
/product="u0002jb"
/protein_id="AAAS0941.1"
/db_xref="GI:560520"
/translacion="MGLVWRSRTSLVGLIGLVRVASFAAQLFYRPSDAVAEEVHKW
YGNLWTKTYMGINCWKSVDMMNYQEILSELOPSLVIEFGTRYGGSAYVAFANMR
QIGQPFKVLTVDNSHKALDPRARREPDLFVSESSSTDPAIAEQRLKNEYFGKIFAI
LDSHSMNVHLAEMKLLRPLLSAGDYLVDVESDINGHPVLPFGFGPGYEAEIAYBDEF
PNDYKDAERENKFGWTSAPNGFLIRN"
12604..13995
/codon_start=1
/transl_table=11
/product="u0002ka"
/protein_id="AAAS0940.1"
/db_xref="GI:560519"
/translacion="MEETSAGDPDPDAGTSTAPNAPEPVARQRILFVGEAATLAH
VREFVLARSIDPSRYEHPACDPRENKLLGLPLPPHPIHTVPSEVLLKIAOGRLF
YNTFLRKTYAADRKIINEIADPVVVDGNRLSLVSARLAGIPYIAIANATWSPARR
RFPDPDVPTRFFRGVRPVSIILYRLRYLFALYCLPLNWLRRKHLSLGLWDLCKRIFT
DGDYTLADVPELVTYNTLPANHYLGPVLMSPDPKPTWWHSLPTDRPIIYATIGSS
GSKNLLQVNALADLPVTIAATAGRNHLKNVPANAFVADLPGEAAAARLAVVLCN
GGSPTQOALAAAGVPVIGLPSMDOHLNMEALERAGAVLLRTERLNTGTEGVAAAQVQ
LSGAEFRQAAERLPPKPSDQTLPSDRTSKVRCGSSAKGCPYRRLANGCPVEHLIRER
VAVNAEVPQRLS"
complement(14306..15133)
/codon_start=1
/transl_table=11
/product="u0002kc"
/protein_id="AAAS0938.1"
/db_xref="GI:560517"
/translacion="MVQTKRYAGLTANTKKVAAWAPMFSIIIPTLNVAAYLPACLDS
IARCTCGDFELVLDGGSDBETLDIANIFAPNLGERLIIRHDTQGVVDANRNGVDLA
TKENILFLGADDSLYEARGLTARVAATFGEHPSDLVYGDVIMRSTNFRMGCAFOLDRL
LFKNNIQAIFVYRGLFGTIGPVNARVLAVDWDFNTRCFSPNALVRYHVVVVASV
NEFGGLSNTIVDEKFLKRLPMSTFLGRLVILVVRPKVISRAMWRTVISWRRR"
complement(15203..15934)
/codon_start=1
/transl_table=11
/product="u0002kb"
/protein_id="AAAS0944.1"
/db_xref="GI:560523"
/translacion="WTSTKVEDRVTAAVLGAIGHALALTASMTWILWALILGFALSA
VVOAVRRESTITVLLGDDRPRTLATATGLGAASSCSVAVALASLFRKGANFTAAAM
APEIGSTNLVVELGIILALLMGWQFTAAEFVGGPMILVLAFLRFLFGARLIDAARE
QAERGLAGSMEGHAAMDSIKREGSFWRRLSLPGFTSIASHVFMELAILRDLILGL
LIRAGIAAWPESFQWQFFFLANHPAWSAVMGPIIGPIVAIVSFVCSIGNVPLAAVLWN
GIGIFGGVIAFIFADLLIPLINIRYKTYGARMMLVLGLTFYASVMVAGYITELLEFGT
TNLIPORSASTMTAEISWNTYTLNVLFLVIAAALVVRFTTSGGLPMLRMWGGSPDA
PHLHDDRHDDHLGH"
9306..10433
/codon_start=1
/transl_table=11
/product="u0002ja"
/protein_id="AAAS0943.1"
/db_xref="GI:560522"
/translacion="WRASRLPALLPSLGPPLWLSRSALHHIPGCGIYTVGNDRAVV
PLNVSPHSCGFPFRVRSVYATASRSGGPPVASEVRGDAISTTPDAAPGLAARRR
RILFVAVTLAHVVRFPALAQSLDPSRYEHPACDPYNNQLLGLPLPRHHAHTIPS
ERFNGNLTQGRFYAMRTLRKYVEADRLVDEIAPDLVVGDLRISLSVSRALAGIPIA
IANAYWSPYAORRFPPLDVIWTRLFGVRLVKLLYRLRPLLPALQCMPLNWMVRRHGL
SSLGNLCRIETDGDHTLYADVPELMTYDLPANHEVILGPVLMSPAGPPTPWDSLPT
DRPIVATYLTGSGGRNLLQLVLNLAELPVTVIAATAAATV"
10439..10846
/codon_start=1
/transl_table=11
/product="u0002h"
/protein_id="AAAS0942.1"
/db_xref="GI:560521"
/translacion="MPANAFVADYLPGEAAAARSAVVVNCNGSLTTOQALVAGVPVIG
VAGNLDQHLNMEAVRACAGVLLRTERLKSQRVAGAVMQVISRSEYPTSRATRCLR
TRPWIIPAAECGCSAPHAGEPAPDMVGQLVTS"
11766..12503
/codon_start=1
/transl_table=11
/product="u0002jb"
/protein_id="AAAS0941.1"
/db_xref="GI:560520"
/translacion="MGLVWRSRTSLVGLIGLVRVASFAAQLFYRPSDAVAEEVHKW
YGNLWTKTYMGINCWKSVDMMNYQEILSELOPSLVIEFGTRYGGSAYVAFANMR
QIGQPFKVLTVDNSHKALDPRARREPDLFVSESSSTDPAIAEQRLKNEYFGKIFAI
LDSHSMNVHLAEMKLLRPLLSAGDYLVDVESDINGHPVLPFGFGPGYEAEIAYBDEF
PNDYKDAERENKFGWTSAPNGFLIRN"
12604..13995
/codon_start=1
/transl_table=11
/product="u0002ka"
/protein_id="AAAS0940.1"
/db_xref="GI:560519"
/translacion="MEETSAGDPDPDAGTSTAPNAPEPVARQRILFVGEAATLAH
VREFVLARSIDPSRYEHPACDPRENKLLGLPLPPHPIHTVPSEVLLKIAOGRLF
YNTFLRKTYAADRKIINEIADPVVVDGNRLSLVSARLAGIPYIAIANATWSPARR
RFPDPDVPTRFFRGVRPVSIILYRLRYLFALYCLPLNWLRRKHLSLGLWDLCKRIFT
DGDYTLADVPELVTYNTLPANHYLGPVLMSPDPKPTWWHSLPTDRPIIYATIGSS
GSKNLLQVNALADLPVTIAATAGRNHLKNVPANAFVADLPGEAAAARLAVVLCN
GGSPTQOALAAAGVPVIGLPSMDOHLNMEALERAGAVLLRTERLNTGTEGVAAAQVQ
LSGAEFRQAAERLPPKPSDQTLPSDRTSKVRCGSSAKGCPYRRLANGCPVEHLIRER
VAVNAEVPQRLS"
complement(14306..15133)
/codon_start=1
/transl_table=11
/product="u0002kc"
/protein_id="AAAS0938.1"
/db_xref="GI:560517"
/translacion="MVQTKRYAGLTANTKKVAAWAPMFSIIIPTLNVAAYLPACLDS
IARCTCGDFELVLDGGSDBETLDIANIFAPNLGERLIIRHDTQGVVDANRNGVDLA
TKENILFLGADDSLYEARGLTARVAATFGEHPSDLVYGDVIMRSTNFRMGCAFOLDRL
LFKNNIQAIFVYRGLFGTIGPVNARVLAVDWDFNTRCFSPNALVRYHVVVVASV
NEFGGLSNTIVDEKFLKRLPMSTFLGRLVILVVRPKVISRAMWRTVISWRRR"
complement(15203..15934)
/codon_start=1
/transl_table=11
/product="u0002kb"
/protein_id="AAAS0944.1"
/db_xref="GI:560523"
/translacion="WTSTKVEDRVTAAVLGAIGHALALTASMTWILWALILGFALSA
VVOAVRRESTITVLLGDDRPRTLATATGLGAASSCSVAVALASLFRKGANFTAAAM
APEIGSTNLVVELGIILALLMGWQFTAAEFVGGPMILVLAFLRFLFGARLIDAARE
QAERGLAGSMEGHAAMDSIKREGSFWRRLSLPGFTSIASHVFMELAILRDLILGL
LIRAGIAAWPESFQWQFFFLANHPAWSAVMGPIIGPIVAIVSFVCSIGNVPLAAVLWN
GIGIFGGVIAFIFADLLIPLINIRYKTYGARMMLVLGLTFYASVMVAGYITELLEFGT
TNLIPORSASTMTAEISWNTYTLNVLFLVIAAALVVRFTTSGGLPMLRMWGGSPDA
PHLHDDRHDDHLGH"
9306..10433
/codon_start=1
/transl_table=11
/product="u0002ja"
/protein_id="AAAS0943.1"
/db_xref="GI:560522"
/translacion="WRASRLPALLPSLGPPLWLSRSALHHIPGCGIYTVGNDRAVV
PLNVSPHSCGFPFRVRSVYATASRSGGPPVASEVRGDAISTTPDAAPGLAARRR
RILFVAVTLAHVVRFPALAQSLDPSRYEHPACDPYNNQLLGLPLPRHHAHTIPS
ERFNGNLTQGRFYAMRTLRKYVEADRLVDEIAPDLVVGDLRISLSVSRALAGIPIA
IANAYWSPYAORRFPPLDVIWTRLFGVRLVKLLYRLRPLLPALQCMPLNWMVRRHGL
SSLGNLCRIETDGDHTLYADVPELMTYDLPANHEVILGPVLMSPAGPPTPWDSLPT
DRPIVATYLTGSGGRNLLQLVLNLAELPVTVIAATAAATV"
10439..10846
/codon_start=1
/transl_table=11
/product="u0002h"
/protein_id="AAAS0942.1"
/db_xref="GI:560521"
/translacion="MPANAFVADYLPGEAAAARSAVVVNCNGSLTTOQALVAGVPVIG
VAGNLDQHLNMEAVRACAGVLLRTERLKSQRVAGAVMQVISRSEYPTSRATRCLR
TRPWIIPAAECGCSAPHAGEPAPDMVGQLVTS"
11766..12503
/codon_start=1
/transl_table=11
/product="u0002jb"
/protein_id="AAAS0941.1"
/db_xref="GI:560520"
/translacion="MGLVWRSRTSLVGLIGLVRVASFAAQLFYRPSDAVAEEVHKW
YGNLWTKTYMGINCWKSVDMMNYQEILSELOPSLVIEFGTRYGGSAYVAFANMR
QIGQPFKVLTVDNSHKALDPRARREPDLFVSESSSTDPAIAEQRLKNEYFGKIFAI
LDSHSMNVHLAEMKLLRPLLSAGDYLVDVESDINGHPVLPFGFGPGYEAEIAYBDEF
PNDYKDAERENKFGWTSAPNGFLIRN"
12604..13995
/codon_start=1
/transl_table=11
/product="u0002ka"
/protein_id="AAAS0940.1"
/db_xref="GI:560519"
/translacion="MEETSAGDPDPDAGTSTAPNAPEPVARQRILFVGEAATLAH
VREFVLARSIDPSRYEHPACDPRENKLLGLPLPPHPIHTVPSEVLLKIAOGRLF
YNTFLRKTYAADRKIINEIADPVVVDGNRLSLVSARLAGIPYIAIANATWSPARR
RFPDPDVPTRFFRGVRPVSIILYRLRYLFALYCLPLNWLRRKHLSLGLWDLCKRIFT
DGDYTLADVPELVTYNTLPANHYLGPVLMSPDPKPTWWHSLPTDRPIIYATIGSS
GSKNLLQVNALADLPVTIAATAGRNHLKNVPANAFVADLPGEAAAARLAVVLCN
GGSPTQOALAAAGVPVIGLPSMDOHLNMEALERAGAVLLRTERLNTGTEGVAAAQVQ
LSGAEFRQAAERLPPKPSDQTLPSDRTSKVRCGSSAKGCPYRRLANGCPVEHLIRER
VAVNAEVPQRLS"
complement(14306..15133)
/codon_start=1
/transl_table=11
/product="u0002kc"
/protein_id="AAAS0938.1"
/db_xref="GI:560517"
/translacion="MVQTKRYAGLTANTKKVAAWAPMFSIIIPTLNVAAYLPACLDS
IARCTCGDFELVLDGGSDBETLDIANIFAPNLGERLIIRHDTQGVVDANRNGVDLA
TKENILFLGADDSLYEARGLTARVAATFGEHPSDLVYGDVIMRSTNFRMGCAFOLDRL
LFKNNIQAIFVYRGLFGTIGPVNARVLAVDWDFNTRCFSPNALVRYHVVVVASV
NEFGGLSNTIVDEKFLKRLPMSTFLGRLVILVVRPKVISRAMWRTVISWRRR"
complement(15203..15934)
/codon_start=1
/transl_table=11
/product="u0002kb"
/protein_id="AAAS0944.1"
/db_xref="GI:560523"
/translacion="WTSTKVEDRVTAAVLGAIGHALALTASMTWILWALILGFALSA
VVOAVRRESTITVLLGDDRPRTLATATGLGAASSCSVAVALASLFRKGANFTAAAM
APEIGSTNLVVELGIILALLMGWQFTAAEFVGGPMILVLAFLRFLFGARLIDAARE
QAERGLAGSMEGHAAMDSIKREGSFWRRLSLPGFTSIASHVFMELAILRDLILGL
LIRAGIAAWPESFQWQFFFLANHPAWSAVMGPIIGPIVAIVSFVCSIGNVPLAAVLWN
GIGIFGGVIAFIFADLLIPLINIRYKTYGARMMLVLGLTFYASVMVAGYITELLEFGT
TNLIPORSASTMTAEISWNTYTLNVLFLVIAAALVVRFTTSGGLPMLRMWGGSPDA
PHLHDDRHDDHLGH"
9306..10433
/codon_start=1
/transl_table=11
/product="u0002ja"
/protein_id="AAAS0943.1"
/db_xref="GI:560522"
/translacion="WRASRLPALLPSLGPPLWLSRSALHHIPGCGIYTVGNDRAVV
PLNVSPHSCGFPFRVRSVYATASRSGGPPVASEVRGDAISTTPDAAPGLAARRR
RILFVAVTLAHVVRFPALAQSLDPSRYEHPACDPYNNQLLGLPLPRHHAHTIPS
ERFNGNLTQGRFYAMRTLRKYVEADRLVDEIAPDLVVGDLRISLSVSRALAGIPIA

```

/protein_id="AAA50939.1"
/db_xref="GI:560518"
/translation="MKSLKLFARFARSAAFEVSRYSERDLKHQFVKQLKSRVDVVF
DVANGGYAAGLRRAAYKGRIVPEPLSGPFTILESASTDPLWDCRQHALGSDGT
VTINTAGAGSSSVLPMLKSHQNAFPFVNTVGTQEASIHRLDSVAPEFLGMNGVAF
KVDVQGFQKVLGAGKSTIDHCVGMQLELSFLPLYEGGMLIPEALDIVSLGFTLTG
LLPCFIDANNRMLQADGIFPREDD"
17208..17933
/codon_start=1
/transl_table=11
/product="u00021"
/protein_id="AAA50937.1"
/db_xref="GI:560516"

CDS

Query Match 43.2%; Score 346; DB 1; Length 31175;
Best Local Similarity 71.0%; Pred. No. 3.5e-63;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;
QY 1 ATGACTGCGCAGTGTCTCGATAATTATCCCTACCTTCAATGACAGCGGTGACGCTGCAA 60
DB 15076 ATGGCGGACCAATGTTTTCGATCATATCCACCTTGAACGTGGCTGCGGTATTGCT 15017
QY 61 GCCTGCCTCGGAAGCATCGTCGGGACACCTACCGGGAAGTGGAAAGTGGTCTCTTGTGAC 120
DB 15016 GCCTGCCTCGACAGCATCGCCCGTCAGACCTGCGGTGACTTCGAGCTGTTACTGTCGAC 14957
QY 121 GCGGTTTCGACCGATCGGACCTCGACATCGGGAACAGTTTCGCGCCCGAACTCGGCTCG 180
DB 14956 GCGGCTCGACGGACGAAACCTCGACATCGCAACATTTTTCGCCCCCAACCTCGCGAG 14897
QY 181 GCACTGGTGTTCACAGCGGGCCCGATGATGGCCCTTACGAGCGCATGAACCGCGGCTC 240
DB 14896 CGTTTGATCATTCATCGGACACCGACGAGCGGCTTACGACGCGCATGAACCGCGGCTG 14837
QY 241 GCGTAGCCACAGGCGAATGGTACTTTTTTATAGCGCGCGACGACACCCCTCTACGAACCA 300
DB 14836 GACCTGGCCACGGAAACGTGGTGTCTTTCTGGCGCGGACGACAGCTGTACGAGCT 14777
QY 301 ACACGTGGCCAGGTAGCGCTTTTCTCGGACCATGCGGCAAGCCATCTTGTCTAT 360
DB 14776 GACACCTGGCGGGTGGCGGCTTCAATGGCGAACAACGAGCCAGCGATCTGGTATAT 14717
QY 361 GCGGATGTTGATGCTTCAACGAAAGCGGACATGCGGACCTTTCGACCTCGACCGC 420
DB 14716 GCGGACGTGATCATGCGCTCAACCAATTCGCTGGGGTGGCGCTTCGACCTCGACCGT 14657
QY 421 CTCCTATTGAGACGAATTTGTGCCCAATTCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
DB 14656 CTGTTGTTCAAGCGCAACATCTGCCATCAGCGCATCTTCTACCGCGCGGACTCTTCGGC 14597
QY 481 GGCATCGGCGCTTACACCTCGCTACCGATCTGGGCGGACTGGGACTTCAATATTGCG 540
DB 14596 ACCATCGTCCCTACAACCTCGCTACCGGCTCTGGCGGACTTGGGACTTCAATATTGCG 14537
QY 541 TGCTTCTCAACCGCGCTGATTACCGCTACATGGAGCTGCTGATTTCCGAATACAAAC 600
DB 14536 TGCTTTTCAACCGAGCGCTGCTACCGCTACATGACGCTGCTGTTGCAAGTACAAAC 14477
QY 601 GACATGACCGGCTTCAGCATGAGGAGGGGACTGATAAGAGTTTCAGAAAAACGGCTGCA 660
DB 14476 GAATTCGGCGGCTCAGCAATACGATCG---TCGACAAGGAGTTTTCGAGCGGCTGCGC 14420
QY 661 ATGTAC 666
DB 14419 ATGTCC 14414

RESULT 12
MSGY2/c MSGY2 31176 bp DNA linear BCT 03-DEC-1996
LOCUS Mycobacterium tuberculosis sequence from clone y2.
DEFINITION
ACCESSION AD000009
VERSION AD000009.1 GI:1702967

KEYWORDS
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE 1 (bases 1 to 31176)
AUTHORS Du, L.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100
Beaver Street, Waltham, MA, USA, 02154 du@cric.com
COMMENT GSDb:S:1004709.
FEATURES
Location/Qualifiers
source 1..31176
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"
/clone="y2"

ORIGIN

Query Match 43.2%; Score 346; DB 1; Length 31176;
Best Local Similarity 71.0%; Pred. No. 3.5e-63;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;
QY 1 ATGACTGCGCAGTGTCTCGATAATTATCCCTACCTTCAATGACAGCGGTGACGCTGCAA 60
DB 15076 ATGGCGGACCAATGTTTTCGATCATATCCACCTTGAACGTGGCTGCGGTATTGCT 15017
QY 61 GCCTGCCTCGGAAGCATCGTCGGGACACCTACCGGGAAGTGGAAAGTGGTCTCTTGTGAC 120
DB 15016 GCCTGCCTCGACAGCATCGCCCGTCAGACCTGCGGTGACTTCGAGCTGTTACTGTCGAC 14957
QY 121 GCGGTTTCGACCGATCGGACCTCGACATCGGGAACAGTTTCGCGCCCGAACTCGGCTCG 180
DB 14956 GCGGCTCGACGGACGAAACCTCGACATCGCAACATTTTTCGCCCCCAACCTCGCGAG 14897
QY 181 CGACTGGTGTTCACAGCGGGCCCGATGATGGCCCTTACGAGCGCATGAACCGCGGCTC 240
DB 14896 CGTTTGATCATTCATCGGACACCGACGAGCGGCTTACGACGCGCATGAACCGCGGCTG 14837
QY 241 GCGTAGCCACAGGCGAATGGTACTTTTTTATAGCGCGCGACGACACCCCTCTACGAACCA 300
DB 14836 GACCTGGCCACGGAAACGTGGTGTCTTTCTGGCGCGGACGACAGCTGTACGAGCT 14777
QY 301 ACACGTGGCCAGGTAGCGCTTTTCTCGGACCATGCGGCAAGCCATCTTGTCTAT 360
DB 14776 GACACCTGGCGGGTGGCGGCTTCAATGGCGAACAACGAGCCAGCGATCTGGTATAT 14717
QY 361 GCGGATGTTGATGCTTCAACGAAAGCGGACATGCGGACCTTTCGACCTCGACCGC 420
DB 14716 GCGGACGTGATCATGCGCTCAACCAATTCGCTGGGGTGGCGCTTCGACCTCGACCGT 14657
QY 421 CTCCTATTGAGACGAATTTGTGCCCAATTCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
DB 14656 CTGTTGTTCAAGCGCAACATCTGCCATCAGCGCATCTTCTACCGCGCGGACTCTTCGGC 14597
QY 481 GGCATCGGCGCTTACACCTCGCTACCGATCTGGGCGGACTGGGACTTCAATATTGCG 540
DB 14596 ACCATCGTCCCTACAACCTCGCTACCGGCTCTGGCGGACTTGGGACTTCAATATTGCG 14537
QY 541 TGCTTCTCAACCGCGCTGATTACCGCTACATGGAGCTGCTGATTTCCGAATACAAAC 600
DB 14536 TGCTTTTCAACCGAGCGCTGCTACCGCTACATGACGCTGCTGTTGCAAGTACAAAC 14477
QY 601 GACATGACCGGCTTCAGCATGAGGAGGGGACTGATAAGAGTTTCAGAAAAACGGCTGCA 660
DB 14476 GAATTCGGCGGCTCAGCAATACGATCG---TCGACAAGGAGTTTTCGAGCGGCTGCGC 14420
QY 661 ATGTAC 666
DB 14419 ATGTCC 14414

RESULT 13
AR345347/c
LOCUS AR345347 3215 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 1 from patent US 6583266.
ACCESSION AR345347
VERSION AR345347.1 GI:33742001
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3215)
AUTHORS Smith,D.R. and Mao,J.-i.
TITLE Nucleic acid and amino acid sequences relating to mycobacterium tuberculosis and leprae for diagnostics and therapeutics
JOURNAL Patent: US 6583266-A 1 24-JUN-2003;
FEATURES
Location/Qualifiers
1..3215
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 43.2%; Score 346; DB 6; Length 32155;
Best Local Similarity 71.0%; Pred. No. 3.5e-63;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;
QY 1 ATGACTGGCCAGTGTCTCGATAATTATCCCTTCAATGACGGGTGACGCTGCAA 60
Db 15076 ATGGCGGCACCAATGTTTCGATCATCCACCCTTGAAGTGGCTGGGTATTGCT 15017
QY 61 GCCTGCTCGGAAGCATGTCGGGAGACCTACCGGGAAGTGAAGTGTCTTGTGAC 120
Db 15016 GCGTCGCTCGACAGCATGCCCGTCCAGACCTCGGTGATCTTCGAGCTGTGTCGAC 14957
QY 121 GCGGCTTCGACGATCGGACCTCGACATCGGAACAGTTTCGCGCCGGAACCTCGGCTCG 180
Db 14956 GCGGCTTCGACGGAACCTTCGACATCGCAACATTTTCGCCCCCAACCTCGGCGAG 14897
QY 181 CACTGTGTGTTACAGCGGGCCGATGATGCCCCCTACGACGCGCATGAACCGGCGCTC 240
Db 14896 CGGTTGATCATTCGCGACACCGACCGAGCGGTCTACGACGCGCATGAACCGGCGGTG 14837
QY 241 GCGGTAGCCACAGGGAATGGTACTTTTATGCGCGGACGACACCTCTACGAACA 300
Db 14836 GACCTGGCCACCGGAACGTTGCTCTTTCTGGCGCGGACGACGCTGTACGAGCT 14777
QY 301 ACCAGCTTGGCCAGGTAGCCGCTTTCTCGCGACCATGCGGCAAGCATCTGTCTAT 360
Db 14776 GACACCTTGGCGGGTGGCGGCTTCATGCGGAACAGAGCCAGCGATCTGTATAT 14717
QY 361 GCGGATGTGTGATCGTTCGACGAAAGCCGACATGCGGACCTTTTCGACCTCGACCGC 420
Db 14716 GCGGACGTGATCATCGCTCAACCAATTTCCGCTGGGTGGCGCTTCGACCTCGACCGT 14657
QY 421 CTCCTATTGTGACAGAAATTGTGCCAACAATGATCTTTTACCGCGGTGAGCTTTTCAC 480
Db 14656 CTGTTGTTCAAGCGGAACATCTGCCATCAGGCGATCTTCTACCGCGCGGACTCTTTCGGC 14597
QY 481 GGCATCGGCCCTTACAACTCGCTACCGAGTCTCGGCGGACTGGGACTTCAATATTGCG 540
Db 14596 ACCATCGGTCCCTACAACTCGCTACCGGCTCGGCGGACTGGGACTTCAATATTGCG 14537
QY 541 TGCTTCTCCAAACCCGGCGTGAATACCGCTACATGAGCGTGTGATTTCCGGAATACAA 600
Db 14536 TGCTTCTCCAAACCCAGCGCTCGTACCGCTACATGACGCTGTGTTGGAAGTACAA 14477
QY 601 GACATGACCGGCTTCAGATGAGGAGGGGACTGATAAAGATTTCAGAAAACGGCTGCCA 660
Db 14476 GAAATTCGGCGGCTTCAGCAATACGATCG---TCGACAAGGAGTTTTGAAGCGGCTGCG 14420
QY 661 ATGTAC 666
Db 14419 ATGTCC 14414

RESULT 14
AE000516 32
WPCOMMENT

Sequence split into 44 fragments LOCUS AE000516 Accession AE000516

Fragment Name

Begin End
AE000516 00 1 110000
AE000516 01 100001 210000
AE000516 02 200001 310000
AE000516 03 300001 410000
AE000516 04 400001 510000
AE000516 05 500001 610000
AE000516 06 600001 710000
AE000516 07 700001 810000
AE000516 08 800001 910000
AE000516 09 900001 1010000
AE000516 10 100001 1110000
AE000516 11 1100001 1210000
AE000516 12 1200001 1310000
AE000516 13 1300001 1410000
AE000516 14 1400001 1510000
AE000516 15 1500001 1610000
AE000516 16 1600001 1710000
AE000516 17 1700001 1810000
AE000516 18 1800001 1910000
AE000516 19 1900001 2010000
AE000516 20 2000001 2110000
AE000516 21 2100001 2210000
AE000516 22 2200001 2310000
AE000516 23 2300001 2410000
AE000516 24 2400001 2510000
AE000516 25 2500001 2610000
AE000516 26 2600001 2710000
AE000516 27 2700001 2810000
AE000516 28 2800001 2910000
AE000516 29 2900001 3010000
AE000516 30 3000001 3110000
AE000516 31 3100001 3210000
AE000516 32 3200001 3310000
AE000516 33 3300001 3410000
AE000516 34 3400001 3510000
AE000516 35 3500001 3610000
AE000516 36 3600001 3710000
AE000516 37 3700001 3810000
AE000516 38 3800001 3910000
AE000516 39 3900001 4010000
AE000516 40 4000001 4110000
AE000516 41 4100001 4210000
AE000516 42 4200001 4310000
AE000516 43 4300001 4403837

Continuation (33 of 44) of AE000516 from base 3200001 (AE000516 Mycobacterium tuberculosis

Query Match 43.2%; Score 346; DB 1; Length 110000;

Best Local Similarity 71.0%; Pred. No. 3.6e-63;

Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

QY 1 ATGACTGGCCAGTGTCTCGATAATTATCCCTTCAATGACGGGTGACGCTGCAA 60
Db 103847 ATGGCGGCACCAATGTTTCGATCATCCACCCTTGAACGCTGCGGTATTGCT 103906
QY 61 GCCTGCTCGGAAGCATGTCGGGACACCTTACCGGGAAGTGAAGTGGTCTTGTTCGAC 120
Db 103907 GCCTGCTCGACAGCATGCGCCGTGACACCTGCGGTGACTTCGAGCTGTGTTGTCAC 103966
QY 121 GCGGTTTCGACCGCATCGGAACCTTCGACATCGGAACAGTTTTCGCGCCGGAACCTCGGCTCG 180
Db 103967 GCGGCTTCGACCGGACGAAACCTTCGACATCGCAACATTTTCGCCCCCAACCTCGCGCAG 104026
QY 181 CGACTGTGTGTTTCACAGCGGCGCGATGATGGCCCTTACGACGCGCATGAACCGCGGCGTC 240
Db 104027 CGGTGTGATCATTCATCGCGACACCGCAGGCGCTCTACGACGCGCATGAACCGCGGCGTG 104086
QY 241 GCGGTGACCAAGCGGAATGGGTACTTTTTTTTAGCGCGCGACGACACCTCTACGAACCA 300

```
Db 104087 GACCTGGCCACCGGAAGTGTTGCTCTTTCTGGCGCGGACGACAGCCTGTACGAGCT 104146
Qy 301 ACCAGCTGGCCAGGTAGCCGCTTTCTCGCGGACGATGGCGGAAGCACTTGTCTAT 360
Db 104147 GACACCTTGGCGCGGTGGCCGCTTCAATTGGCGAACAAGAGCCGAGGATCTGGTATAT 104206
Qy 361 GCGGATGTTGTATCGTTTCGACGAAAGCGGCATGCCGACCTTTCGACCTCGACCGC 420
Db 104207 GCGGACGTGATCATGGCTCAACCAATTTCCGCTGGGTGGCGCTTGCACCTCGACCGT 104266
Qy 421 CTCCTATTGTAGACGAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
Db 104267 CTGTTGTTCAAGCGCAACATCTGCCATCAGGCGATCTTCTACCGCGCGGACTCTTCGGC 104326
Qy 481 GGCATCGGCCCTTACAACCTGGCTACCGAGTCTGGCGGAGTGGGACTTCAATATTGCG 540
Db 104327 ACCATCGGTCCCTACAACCTCGCTACCGGTCTCGCGGACTTGGGACTTCAATATTGCG 104386
Qy 541 TGCCTTCTCAACCCGCGCTGATTACCCGCTACATGGACGTGCTGATTTCCGAATACAAC 600
Db 104387 TGCCTTTTCAACCCAGCGCTGCTACCCGCTACATGCACTGCTGTTGCAAGCTACAAC 104446
Qy 601 GACATGACCGGCTTCAGCATGAGGAGGCGGACTTGATAAAGAGTTTCAGAAAACGGCTGCCA 660
Db 104447 GAATTCGGCGGCTCAGCAATACGATCG---TCGACAGGAGTTTTCGAGCGGCTGCCG 104503
Qy 661 ATGTAC 666
Db 104504 ATGTCC 104509
```

RESULT 15

AE000516_33

WPCOMMENT

Sequence split into 44 fragments LOCUS AE000516 Accession AE000516

Fragment Name	Begin	End
AE000516_00	1	110000
AE000516_01	100001	210000
AE000516_02	200001	310000
AE000516_03	300001	410000
AE000516_04	400001	510000
AE000516_05	500001	610000
AE000516_06	600001	710000
AE000516_07	700001	810000
AE000516_08	800001	910000
AE000516_09	900001	1010000
AE000516_10	1000001	1110000
AE000516_11	1100001	1210000
AE000516_12	1200001	1310000
AE000516_13	1300001	1410000
AE000516_14	1400001	1510000
AE000516_15	1500001	1610000
AE000516_16	1600001	1710000
AE000516_17	1700001	1810000
AE000516_18	1800001	1910000
AE000516_19	1900001	2010000
AE000516_20	2000001	2110000
AE000516_21	2100001	2210000
AE000516_22	2200001	2310000
AE000516_23	2300001	2410000
AE000516_24	2400001	2510000
AE000516_25	2500001	2610000
AE000516_26	2600001	2710000
AE000516_27	2700001	2810000
AE000516_28	2800001	2910000
AE000516_29	2900001	3010000
AE000516_30	3000001	3110000
AE000516_31	3100001	3210000
AE000516_32	3200001	3310000
AE000516_33	3300001	3410000
AE000516_34	3400001	3510000
AE000516_35	3500001	3610000

```
AE000516_36 3600001 3710000
AE000516_37 3700001 3810000
AE000516_38 3800001 3910000
AE000516_39 3900001 4010000
AE000516_40 4000001 4110000
AE000516_41 4100001 4210000
AE000516_42 4200001 4310000
AE000516_43 4300001 4403837
```

Continuation (34 of 44) of AE000516 from base 3300001 (AE000516 Mycobacterium tuberculosis)

```
Query Match 43.2%; Score 346; DB 1; Length 110000;
Best Local Similarity 71.0%; Pred. No. 3.6e-63;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

Qy 1 ATGACTGGCCAGTGTCTCGATAATTATCCCTACCTTCATATGACAGCGGTGACGCTGCAA 60
Db 3847 ATGCGCGACCAATGTTTTCGATCATATCCCCACCTTGAACGTGGCTGCGGTATTGCT 3906
Qy 61 GCCTGCCTCGGAAGCATCGTCGGGCAGACCTTACCGGGAAGTGAAGTGGTCTTGTTCGAC 120
Db 3907 GCCTGCCTCGACAGCATGCGCCGTCAGACCTGCGGTGACTTCGAGCTGGTACTGGTCAC 3966
Qy 121 GCGGTTTCGACCGCATCGAACCTTCGACATCGCAACAGTTTTCGCCCGCGGAACCTCGGTCG 180
Db 3967 GCGGCTCGACGGACGAAACCTTCGACATCGCAACATTTTTCGCCCCCAACCTCGCGGAG 4026
Qy 181 CGACTGTGCTTCACAGGGGCGCGATGATGGCCCTTACGACGCATGAACCGGGCGGTC 240
Db 4027 CGGTTGATCATTTTCGCGACACCGGCGCTTACGACGCATGAACCGGGCGGTG 4086
Qy 241 GCGGTAGCCACAGCGGAATGGGTACTTTTTAGCGCGCGACGACACCTCTACGAACCA 300
Db 4087 GACCTGGCCACCGGAACGTGGTTGCTCTTTCGGGCGGAGACGACGCTGTACGAGCT 4146
Qy 301 ACCAGTTTGGCCAGGTAGCGGCTTTTCTCGGACCACTTCGCGCAAGCCATCTTGTCTAT 360
Db 4147 GACACCTTGGCGGGTGGCGGCTTCATTGGCGAAACAGAGCCAGCGATCTGGTATAT 4206
Qy 361 GCGCATGTTGTAGCGTTCGACGAAAGCGGATGCGCGACCTTTCGACCTCGACCGC 420
Db 4207 GCGCAGCTGATCATGCGCTCAACCAATTTCCGCTGGGGTGGCGCTTCGACCTCGACCGT 4266
Qy 421 CTCCTATTGAGACGAATTTGTGCGCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
Db 4267 CTGTTGTTCAAGCGCAACATCTGCCATCAGCGGATCTTACCGCGCGGACTCTTCGGC 4326
Qy 481 GGCATCGGCCCTTACAACCTGCGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTGCG 540
Db 4327 ACCATCGGTCCCTACAACTCCGCTACCGGTCCTGGCCGACTGGGACTTCAATATTGCG 4386
Qy 541 TGCTTCTCAACCGCGGCTGATTTACCGCTTACATGGAGCTCGTATTTCGAATACAAC 600
Db 4387 TGCTTTTCCAAACCCAGCGCTCGTCAACCGCTTACATGCACTGGTCTGTTCGAAGCTACAAC 4446
Qy 601 GACATGACCGGCTTCAGCATGAGGAGGGAAGTGTATAAAGAGTTTCAGAAAACGGCTGCCA 660
Db 4447 GAATTCGGCGGCTCAGCAATACGATCG---TCGACAGGAGTTTTCGAGCGGCTGCCG 4503
Qy 661 ATGTAC 666
Db 4504 ATGTCC 4509
```

Search completed: March 6, 2005, 21:05:08
Job time : 3865 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2005, 21:05:14 ; Search time 597 Seconds
(without alignments)
7961.493 Million cell updates/sec.

Title: US-10-805-311-23

Perfect score: 801

Sequence: 1 atgactgcccagtggtctc.....aacgaagcgcagaaccgtag 801

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5401638 seqs, 2966923429 residues

Total number of hits satisfying chosen parameters: 10803276

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/US06_PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09D_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	801	100.0	801	18	US-10-805-311-23
2	801	100.0	4435	18	US-10-805-311-4
3	799.4	99.8	801	18	US-10-805-311-21
4	799.4	99.8	7995	18	US-10-805-311-3
5	346	43.2	768	17	US-10-282-122A-26361
6	346	43.2	828	17	US-10-282-122A-28624
7	346	43.2	828	18	US-10-805-311-38
8	69.2	8.6	2334	15	US-10-156-761-5079
9	69.2	8.6	9025608	15	US-10-156-761-1
10	67.4	8.4	2196	15	US-10-156-761-5077
11	60	7.5	900	17	US-10-282-122A-30023

12	57	7.1	2256646	18	US-10-470-565-1	Sequence 1, Appli
13	55.6	6.9	1060	15	US-10-193-002-306	Sequence 306, App
14	55.6	6.9	1060	15	US-10-084-843-311	Sequence 311, App
15	55.4	6.9	1033	17	US-10-282-122A-36601	Sequence 36601, A
16	53.8	6.7	1035	17	US-10-282-122A-39197	Sequence 39197, A
17	53.4	6.7	4512	15	US-10-156-761-2222	Sequence 2222, Ap
18	48	6.0	738	17	US-10-424-599-107538	Sequence 107538,
19	47.8	6.0	1022	15	US-10-193-002-325	Sequence 325, App
20	47.8	6.0	1022	15	US-10-084-843-330	Sequence 330, App
21	45.2	5.6	1827	17	US-10-108-260A-139	Sequence 139, App
22	42.8	5.3	753	15	US-10-156-761-6137	Sequence 6137, Ap
23	42.8	5.3	4725	15	US-10-205-032-17	Sequence 17, Appli
24	42.8	5.3	5859	13	US-10-007-267-7	GENERAL INFORMA
25	42.8	5.3	5859	13	US-10-007-267-7	GENERAL INFORMA
26	42.8	5.3	5859	17	US-10-096-129-1	Sequence 1, Appli
27	42.8	5.3	5859	17	US-10-654-528-1	Sequence 1, Appli
28	42.8	5.3	5859	17	US-10-654-528-7	Sequence 7, Appli
29	42.8	5.3	60196	15	US-10-205-032-1	Sequence 1, Appli
30	42.8	5.3	9025608	15	US-10-156-761-1	Sequence 1, Appli
31	42.6	5.3	1326	17	US-10-282-122A-23449	Sequence 23449, A
32	42.4	5.3	1139	18	US-10-425-115-78950	Sequence 78950, A
33	42.4	5.3	1143	15	US-10-156-761-823	Sequence 823, App
34	42.2	5.3	2664	18	US-10-437-963-61100	Sequence 61100, A
35	41.4	5.2	522	17	US-10-424-599-73486	Sequence 73486, A
36	40.8	5.1	320	16	US-10-259-678-467	Sequence 467, App
37	40.8	5.1	882	17	US-10-282-122A-33586	Sequence 33586, A
38	40.8	5.1	1836	10	US-09-953-348-52	Sequence 52, Appl
39	40.8	5.1	1836	15	US-10-267-255-52	Sequence 52, Appl
40	40.8	5.1	53500	10	US-09-953-348-76	Sequence 76, Appl
41	40.8	5.1	53500	15	US-10-267-255-76	Sequence 76, Appl
42	40.6	5.1	2073	17	US-10-282-122A-11384	Sequence 11384, A
43	40.6	5.1	2427	18	US-10-437-963-69728	Sequence 69728, A
44	40.4	5.0	1443	15	US-10-156-761-4460	Sequence 4460, Ap
45	40.4	5.0	2811	10	US-09-953-348-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1

US-10-805-311-23
; Sequence 23, Application US/10805311
; Publication No. US20040260078A1
; GENERAL INFORMATION:
; APPLICANT: Hermon-Taylor, John
; APPLICANT: Doran, Tim
; APPLICANT: Millar, Douglas
; APPLICANT: Tizard, Mark
; APPLICANT: Loughlin, Mark
; APPLICANT: Sumar, Nazira
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY
; TITLE OF INVENTION: TARGETS FOR CHEMOTHERAPY
; FILE REFERENCE: 117-260
; CURRENT APPLICATION NUMBER: US/10/805,311
; CURRENT FILING DATE: 2004-03-22
; PRIOR APPLICATION NUMBER: US/09/705,911
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: US/09/091,538
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: PCT/GB96/03221
; PRIOR FILING DATE: 1996-12-23
; PRIOR APPLICATION NUMBER: GB 9526178.0
; PRIOR FILING DATE: 1995-12-21
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Mycobacterium
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(798)

```
US-10-805-311-23
Query Match      100.0%; Score 801; DB 18; Length 801;
Best Local Similarity 100.0%; Pred. No. 7.6e-255; Indels 0; Gaps 0;
Matches 801; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACTGGCCAGAGTCTTCGATATTAATCCCTACCTTCAATGACAGCGGTGACGCTGCAA 60
DB 1 ATGACTGGCCAGAGTCTTCGATATTAATCCCTACCTTCAATGACAGCGGTGACGCTGCAA 60
QY 61 GCCTGCTCGGAAGCATCGTCGGGAGACCTACCGGGAAGTGGAGTGGTCTTGTGCGAC 120
DB 61 GCCTGCTCGGAAGCATCGTCGGGAGACCTACCGGGAAGTGGAGTGGTCTTGTGCGAC 120
QY 121 GCGGTTTCGACCGATCGACCTCGACATCGGAACAGTTCCTCGCCCGGAACCTCGGCTCG 180
DB 121 GCGGTTTCGACCGATCGACCTCGACATCGGAACAGTTCCTCGCCCGGAACCTCGGCTCG 180
QY 181 CGACTGGTCTGTTACAGCGGGCCCGATGATGSCCCCTACGACGCGCATGAACCGGGCGTC 240
DB 181 CGACTGGTCTGTTACAGCGGGCCCGATGATGSCCCCTACGACGCGCATGAACCGGGCGTC 240
QY 241 GCGGTAGCCACAGCGCAATGGGTACTTTTTTATGGCGCCGACGACACCTCTACGAACCA 300
DB 241 GCGGTAGCCACAGCGCAATGGGTACTTTTTTATGGCGCCGACGACACCTCTACGAACCA 300
QY 301 ACCAGTTGGCCAGGTAGCGCTTTCTCGGAGCAATGCGGCAAGCATCTTGTCTAT 360
DB 301 ACCAGTTGGCCAGGTAGCGCTTTCTCGGAGCAATGCGGCAAGCATCTTGTCTAT 360
QY 361 GCGGATGTTGTATGTCGTTTCAGCAAGAAAGCGGACCTTCCTCGACTCGACCGC 420
DB 361 GCGGATGTTGTATGTCGTTTCAGCAAGAAAGCGGACCTTCCTCGACTCGACCGC 420
QY 421 CTCCTATTGAGACGAATTTGTGCCAACAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
DB 421 CTCCTATTGAGACGAATTTGTGCCAACAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
QY 481 GGCATCGGCCCTTACAACTCGGCTTACCGAGTCTGGGCGGACTGGGACTTCAATATTGCG 540
DB 481 GGCATCGGCCCTTACAACTCGGCTTACCGAGTCTGGGCGGACTGGGACTTCAATATTGCG 540
QY 541 TGCCTTCTCAACCCGGCGTANTTACCGCTACATGGAGCTGCTGATTTCCGAATACAA 600
DB 541 TGCCTTCTCAACCCGGCGTANTTACCGCTACATGGAGCTGCTGATTTCCGAATACAA 600
QY 601 GACATGACCGGCTTCAGCATGAGGCGAGGAGCTGATAAAGAGTTTCAGAAAACGGCTGCCA 660
DB 601 GACATGACCGGCTTCAGCATGAGGCGAGGAGCTGATAAAGAGTTTCAGAAAACGGCTGCCA 660
QY 661 ATGTACTTCTGGGTTGCAAGGTGGGAGACTTTCAGGCGGATGCTGGCGTTTGTGAAGAC 720
DB 661 ATGTACTTCTGGGTTGCAAGGTGGGAGACTTTCAGGCGGATGCTGGCGTTTGTGAAGAC 720
QY 721 AAGGAGAAATCGCGCTCGCTTGGTACGCGGTTGATAAGGGTTAAGCGCTCTCCAAA 780
DB 721 AAGGAGAAATCGCGCTTGGCTTGGCTACGCGGTTGATAAGGGTTAAGCGCTCTCCAAA 780
QY 781 GAACGAAGCGCAGAACCGGTAG 801
DB 781 GAACGAAGCGCAGAACCGGTAG 801

RESULT 2
US-10-805-311-4
; Sequence 4, Application US/10805311
; Publication No. US20040260078A1
; GENERAL INFORMATION:
; APPLICANT: Hermon-Taylor, John
; APPLICANT: Doran, Tim
; APPLICANT: Millar, Douglas
; APPLICANT: Tizard, Mark
; APPLICANT: Loughlin, Mark
```


Db 3995 ATGTAATCTCGGTTGCGAGTGGGAGACTTCAGCGCGCATGCTGGCGTTTGTGAAGAC 4054
Qy 721 AAGGAGAAATCGCGTCTGCGCTTGGGTAGCGCGTTGATAAGGGTTAAGCCCGTCTCCAAA 780
Db 4055 AAGGAGAAATCGCGTCTGCGCTTGGGTAGCGCGTTGATAAGGGTTAAGCCCGTCTCCAAA 4114
Qy 781 GAACGAAGCGGAGAACCGTAG 801
Db 4115 GAACGAAGCGGAGAACCGTAG 4135

RESULT 3
US-10-805-311-21
; Sequence 21, Application US/10805311
; Publication No. US20040260078A1
; GENERAL INFORMATION:
; APPLICANT: Hermon-Taylor, John
; APPLICANT: Doran, Tim
; APPLICANT: Millar, Douglas
; APPLICANT: Tizard, Mark
; APPLICANT: Loughlin, Mark
; APPLICANT: Sumar, Nazira
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC
; TITLE OF INVENTION: MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND
; TITLE OF INVENTION: TARGETS FOR CHEMOTHERAPY
; FILE REFERENCE: 117-260
; CURRENT APPLICATION NUMBER: US/10/805,311
; CURRENT FILING DATE: 2004-03-22
; PRIOR APPLICATION NUMBER: US/09/705,911
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: US/09/091,538
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: PCT/GB96/03221
; PRIOR FILING DATE: 1996-12-23
; PRIOR APPLICATION NUMBER: GB 9526178.0
; PRIOR FILING DATE: 1995-12-21
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 21
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Mycobacterium
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(798)
US-10-805-311-21

Query Match 99.8%; Score 799.4; DB 18; Length 801;
Best Local Similarity 99.9%; Pred. No. 2.6e-254;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGACTGCGCCAGTGTCTCGATAATTATCCCTACCTTCAATGACGCGGTGACGCTGCAA 60
Db 1 ATGACTGCGCCAGTGTCTCGATAATTATCCCTACCTTCAATGACGCGGTGACGCTGCAA 60
Qy 61 GCCTGCTCGGAAGCATGCTGCGGCGAGACCTACCGGGGAAGTGGAGTGGTCTTGTGAC 120
Db 61 GCCTGCTCGGAAGCATGCTGCGGCGAGACCTACCGGGGAAGTGGAGTGGTCTTGTGAC 120
Qy 121 GCGGTTGACGATCGGACCTCGACATCGGGAACAGTTTCGCGCCCGGAACCTCGGCTCG 180
Db 121 GCGGTTGACGATCGGACCTCGACATCGGGAACAGTTTCGCGCCCGGAACCTCGGCTCG 180
Qy 181 CGACTGGTCTGTTACAGCGGGCGCCATGATGCGCCCTACGACGCGCATGAACCGGCGGTC 240
Db 181 CGACTGGTCTGTTACAGCGGGCGCCATGATGCGCCCTACGACGCGCATGAACCGGCGGTC 240
Qy 241 GCGGTAGCCACAGGCGAATGGGTACTTTTTAGCGCGCGGACGACACCTCTACGAACCA 300
Db 241 GCGGTAGCCACAGGCGAATGGGTACTTTTTAGCGCGCGGACGACACCTCTACGAACCA 300
Qy 301 ACCAGTTGGCCAGGTAGCGCGTTTCTCGGCGAACCATGCGGCAAGCCATCTTGTCTAT 360
Db 301 ACCAGTTGGCCAGGTAGCGCGTTTCTCGGCGAACCATGCGGCAAGCCATCTTGTCTAT 360

Db 301 ACCAGTTGGCCAGGTAGCGCGTTTCTCGGCGAACCATGCGGCAAGCCATCTTGTCTAT 360
Qy 361 GCGCATGTTGTGATGCGTTTCGACGAAAAAGCGCGCATGCGGACCTTTTCGACCTCGACCGC 420
Db 361 GCGCATGTTGTGATGCGTTTCGACGAAAAAGCGCGCATGCGGACCTTTTCGACCTCGACCGC 420
Qy 421 CTCTTATTGTGAGCGAATTTGTGCGCAACCAATCGATCTTTTACCGCGTGAGCTTTTCGAC 480
Db 421 CTCTTATTGTGAGCGAATTTGTGCGCAACCAATCGATCTTTTACCGCGTGAGCTTTTCGAC 480
Qy 481 GGCATGCGCCCTTACACCTCGGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTTCG 540
Db 481 GGCATGCGCCCTTACACCTCGGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTTCG 540
Qy 541 TGTCTTCTCCAAACCGCGCGCTGATTACCGCTACATGAGCGTCTGATTTCCGAAATACAAC 600
Db 541 TGTCTTCTCCAAACCGCGCGCTGATTACCGCTACATGAGCGTCTGATTTCCGAAATACAAC 600
Qy 601 GACATGACCGGCTTCAGCATGAGGAGGGGACTGATAAAGAGTTTCAGAAAAACGGCTGCCA 660
Db 601 GACATGACCGGCTTCAGCATGAGGAGGGGACTGATAAAGAGTTTCAGAAAAACGGCTGCCA 660
Qy 661 ATGTACTTCTGGGTTGAGGCTGGGAGTTCGAGGCGCATGCTGGCGTTTGTGAAGAC 720
Db 661 ATGTACTTCTGGGTTGAGGCTGGGAGTTCGAGGCGCATGCTGGCGTTTGTGAAGAC 720
Qy 721 AAGGAGAAATCGCGTCTGCGCTTGGTACGCGGTTGATAAGGGTTAAGCGCGTCTCCAAA 780
Db 721 AAGGAGAAATCGCGTCTGCGCTTGGTACGCGGTTGATAAGGGTTAAGCGCGTCTCCAAA 780
Qy 781 GAACGAAGCGGAGAACCGTAG 801
Db 781 GAACGAAGCGGAGAACCGTAG 801

RESULT 4
US-10-805-311-3
; Sequence 3, Application US/10805311
; Publication No. US20040260078A1
; GENERAL INFORMATION:
; APPLICANT: Hermon-Taylor, John
; APPLICANT: Doran, Tim
; APPLICANT: Millar, Douglas
; APPLICANT: Tizard, Mark
; APPLICANT: Loughlin, Mark
; APPLICANT: Sumar, Nazira
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC
; TITLE OF INVENTION: MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND
; TITLE OF INVENTION: TARGETS FOR CHEMOTHERAPY
; FILE REFERENCE: 117-260
; CURRENT APPLICATION NUMBER: US/10/805,311
; CURRENT FILING DATE: 2004-03-22
; PRIOR APPLICATION NUMBER: US/09/705,911
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: US/09/091,538
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: PCT/GB96/03221
; PRIOR FILING DATE: 1996-12-23
; PRIOR APPLICATION NUMBER: GB 9526178.0
; PRIOR FILING DATE: 1995-12-21
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 7995
; TYPE: DNA
; ORGANISM: Mycobacterium
US-10-805-311-3

Query Match 99.8%; Score 799.4; DB 18; Length 7995;
Best Local Similarity 99.9%; Pred. No. 5.8e-254;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGACTGCGCCAGTGTCTCGATAATTATCCCTACCTTCAATGACGCGGTGACGCTGCAA 60
Db 1 ATGACTGCGCCAGTGTCTCGATAATTATCCCTACCTTCAATGACGCGGTGACGCTGCAA 60

[illegible]

RESULT 5

```

RES0001.3
US-10-282-122A-26361
; Sequence 26361, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

```



```
; ORGANISM: Mycobacterium
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(825)
US-10-805-311-38

Query Match      43.2%; Score 346; DB 18; Length 828;
Best Local Similarity 71.0%; Pred. No. 6.8e-104;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

QY 1 ATGACTGCGCCAGTGTCTCGATAATTAATCCCTACCTTCAATGACGCGGTGACGCTGCAA 60
Db 58 ATGGCGCACCAATGTTTTCGATCATATCCCACTTGAACTGCGTGGTATTGCT 117
QY 61 GCCTGCTCGGAAGATCTGCGGACACCTACCGGGAAGTGAAGTGTCTTGTTCGAC 120
Db 118 GCCTGCTCGACAGCATCCCGGTGAGACTCGGTGACTTCGAGTGTGTACTGTGTCGAC 177
QY 121 GCGGTTGCGACGATCGGACCTCGACATCGGAACAGTTTCCGCCCGGAACCTCGGCTCG 180
Db 178 GCGGTTGCGACGAGAAACCTCGACATCGCAACATTTTCGCCCGCAACCTCGGCGAG 237
QY 181 CGACTGCTGTTTACAGCGGGCCCGATGATGTCGCCCTACGACGCCATGAACCGCGCGTTC 240
Db 238 CGGTTGATCATTCATCGGACACCGACGAGGCGTCTACGACGCCATGAACCGCGCGGTG 297
QY 241 GCGGTAGCCACAGGCGAATGGGTACTTTTTTTAGGCGCGACGACACCTCTTACGAACA 300
Db 298 GACCTGGCCACCGGAACGTTGGTGTCTTTCTTGGGCGCGGACGACAGCTGTACGAGGT 357
QY 301 ACCAGTTGGCCAGGTAGCGCTTTCTCGGCGACCATGCGGCAAGCCATCTTGTCTAT 360
Db 358 GACACCTTGGCGGGGTGGCGCCCTTCATTTGGCGAACACGAGCCAGCGATCTGTATAT 417
QY 361 GCGGATGTTGTGATCGTTTCGACGAAAGCCGGCATGCCGACCTTTTCGACCTCGACCGC 420
Db 418 GCGGACGTGATCATGGCTCAACCAATTTCCGCTGGGGTGGCGCTTCGACCTCGACCGT 477
QY 421 CTCCTATTGAGACGAAATTTGGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
Db 478 CTGTTGTTCAAGCGCAACATCTGCCATCAGGCGATCTTCTACCGCGCGGACTCTTTCGCG 537
QY 481 GGCATCGGCCCTTACAACTCGCTGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTCG 540
Db 538 ACATCGGTCCTTACAACTCCGCTACCGGTCCTGGCGACTGGGACTTCAATATTCG 597
QY 541 TCGTTCTCCAAACCGCGCGCTGATTACCGCTACATGAGCTGCTGATTTCGGAATACAAAC 600
Db 598 TCGTTTCCAAACCGCGGCTCGTCAACCGCTACATGCACTGCTGTTGCAAGCTTACAAAC 657
QY 601 GACATGACCGGCTTACGATAGGCGAGGGAATGATAAGATTTCAGAAACGGCTGCCA 660
Db 658 GAATTCGGCGGCTCAGCAATACGATCG---TCGAAAGGAGTTTTTGAAGCGGCTGCGG 714
QY 661 ATGTAC 666
Db 715 ATGTCC 720
```

```
RESULT 8
US-10-156-761-5079
; Sequence 5079, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
```

```
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5079
; LENGTH: 2334
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2334)
US-10-156-761-5079

Query Match      8.6%; Score 69.2; DB 15; Length 2334;
Best Local Similarity 52.8%; Pred. No. 6.6e-12;
Matches 149; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 9 GCCAGTGTTCGATAAATATCCCTACCTTCAATGACGCGGTGACGCTGCAAGCCTGCT 68
Db 30 GCCAGGTTTCAGTGTGTCATCGTCCCGCGTACAAGGTTTCAGGCGTACCTGTCATGAGT 89
QY 69 CGGAGCATCTCGGGCAGACCTACCGGGAAGTGAAGTGTCTTGTGCGACGGGTTTC 128
Db 90 CGAATCCGTGTGGAGCAGTCTTCCCGGACCTCGAATGATCGGCGTTCGACGACTGCTC 149
QY 129 GACCATCGGACCTCGACATCGGAAACAGTTTTCGCCCGGAACCTCGGCTCGCGACTGCT 188
Db 150 ACCGACGCTCGGGCGGATCATCGACGAGTTCGCGGCCGTGACACCCGGGTACGCCC 209
QY 189 CGTTCACAGCGGCGCGGATGATGAGCCCTTACGACGCCATGAACCGCGCGGTGCGGTA 248
Db 210 CGTCACCTCGACAGAACACGCGGTCTGGGCGGCGCCGCAACGCGGCCCTTGGCCGAGGC 269
QY 249 CACAGGCGAATGGGTACTTTTTTTAGGCGCGCGACGACACCCCT 290
Db 270 CACGCGCGACTACCTGATCTTCTTGGACGCGCGACGACGCT 311
```

```
RESULT 9
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      8.6%; Score 69.2; DB 15; Length 9025608;
```

Best Local Similarity 52.8%; Pred. No. 1.2e-10;
Matches 149; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

```
QY 9 GCCAGTGTCTCGATAATTTATCCCTACCTTCAATCAGCGGTGACGCTGCAAGCCTGCCT 68
Db 6184421 GCCAGTGTCTCGATAATTTATCCCTACCTTCAATCAGCGGTGACGCTGCAAGCCTGCCT 6184362

QY 69 CGGAAGCATCTGTCGGGCGAGACCTACCGGGAAGTGGTCTTGTGACGGCGGTTTC 128
Db 6184361 CGAATCCGTGCTGGAGCAGTCTCTCCCGACCTCGAATGATCGCGGTGACGACTGCTC 6184302

QY 129 GACCATCGGACCTCGACATCGGAACAGTTTCCGCCCGGAACCTCGGCTCGGACTGGT 188
Db 6184301 ACCCGACGCTCTCGGCGCGATCATCGACGAGTTTCGGGCGCCGTGACACCCGGGTACGGCC 6184242

QY 189 CGTTCACAGCGGCGCGATGATGGCCCTTACGACCCATGAACCGCGGTGCGGCTAGC 248
Db 6184241 CGTCACCTGCCACAGAACAGCGGCTGCGGCGCGCCGCAACGCGGCCCTGCGCCAGGC 6184182

QY 249 CACAGCGGAATGGGTACTTTTTTATAGGCGCCGACGACACCTT 290
Db 6184181 CACCGCGCACTACCTGATCTTCTGAGCGCGCAACGACAGCCT 6184140
```

RESULT 10

US-10-156-761-5077
; Sequence 5077, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5077
; LENGTH: 2196
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2196)
US-10-156-761-5077

Query Match 8.4%; Score 67.4; DB 15; Length 2196;
Best Local Similarity 52.3%; Pred. No. 2.6e-11;
Matches 149; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

```
QY 7 GGCAGCATGTTCTCGATAATTTATCCCTACCTTCAATGACGCGGTGACGCTGCAAGCCTGC 66
Db 1 GTGCCCGCTTCAGCATCATCTGCTCCCGCTTCTTCAAGGTGACGGGTATCTGCGCGAGTGC 60

QY 67 CTCGGAAGCATCTGTCGGGCGAGACCTACCGGGAAGTGGAGTGGTCTTGTGACGGCGGT 126
Db 61 CTCGACTCGGTCTCTGGGCGAGTCTGACCGGATCTGGAAGTATCGCGTTCGACACTGC 120

QY 127 TCAGCCGATCGGACCTCGACATCGGAACAGTTTCCGCCCGGAACCTCGGCTCGGACTG 186
Db 121 TCGCGGAGCGGTCTGGGCGCCATCTTCGACGGGTACGCGGCCGCGACCCGCGGTACGC 180

QY 187 GTGCTTCAGCGGGCGCGATGATGGCCCTTACGACGCGCATGAACCGGGCGGTTCGGCGTA 246
Db 181 GTGCTGCACCTCTCCCGAGAACTGTCGCTTGGGCGCGGCAACGCAACGCGGGAATGCGGCAC 240
```

```
QY 247 GCCACAGCGAATGGGTACTTTTTTATAGCGCCGACGACACCCCTC 291
Db 241 GCCACCGGGGACTACCTTCTTCTCTCGACAGCGACACCCCTC 285
```

RESULT 11

US-10-282-122A-30023
; Sequence 30023, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Truick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30023
; LENGTH: 900
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30023

Query Match 7.5%; Score 60; DB 17; Length 900;
Best Local Similarity 50.0%; Pred. No. 5.3e-09;
Matches 209; Conservative 0; Mismatches 200; Indels 9; Gaps 2;

```
QY 7 GCGCCAGTGTCTCGATAATTTATCCCTACCTTCAATGACGCGGTGACGCTGCAAGCCTGC 66
Db 46 GCGCCGCTGTATCCGTCTAGCGCCCTTCAACCGGGAAGTACTCTGGAAGAGCC 105

QY 67 CTCGGAAGCATCTGTCGGGCGAGACCTACCGGGAAGTGGAGTGGTCTTGTGACGGCGGT 126
Db 106 CTGCGGAGCATCTACGAGCAGGACTACCCGAAATTTTGAAGTGATCATCTGTCGACGAGCT 165

QY 127 TCGACCGATCGGACCTCGACATCGGAACAGTTTCCGCCCGGAACCTCGGCTCGGACTG 186
Db 166 TCCACCGAACACAGCTACGCCATCTGGAGCAGTTGAGAAAGTCCACGGTTTCCAGC-- 223

QY 187 GTCGTTACAGCGGCGCGCATGATGGCCCTTACGACGCCATGAACCGCGGCGGTTCGGCGTA 246
```

Db 224 ---TCTACCGCAGCAGAACAGCGCTGACGCGAGCTGAACTTCGGCTCGACAC 279
Qy 247 GCCACAGCGGAATGGTACTTTTTTAGCGCGCGACGACACCTCTACGAACCAACGACG 306
Db 280 GCCCGCGGCGACTAGCTAGCCACGCGGACCTGGACGACA--TCTGCTCGCGACTCG 336
Qy 307 TTGGCCCGAGTAGCGCTTTTCTCGCGACCATGCGGCAAGCATCTTGTCTATGGCGAT 366
Db 337 TTGACGCTGCGCGCGCTACCTCGACGACACCGGAAAGTGGCTGCTGGCGCCCTG 396
Qy 367 GTTGATGTCGTTCCAGCAAAAGCGCATGCGCGACCTTTTCGACTCGACCGCCTCC 424
Db 397 GTGATCTACATCGACAGCGAGGCCAGGAACCAAGCGGCAATGCGATCC 454

RESULT 12

US-10-470-565-1
; Sequence 1, Application US/10470565
; Publication No. US20040126870A1
; GENERAL INFORMATION:
; APPLICANT: Societe des Produits Nestle S.A.
; TITLE OF INVENTION: NCC2705 - the genome of a Bifidobacterium
; FILE REFERENCE: 80290/WO
; CURRENT APPLICATION NUMBER: US/10/470,565
; CURRENT FILING DATE: 2003-07-29
; PRIOR APPLICATION NUMBER: EP 01102050.0
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2256646
; TYPE: DNA
; ORGANISM: Bifidobacterium longum
US-10-470-565-1

Query Match 7.1%; Score 57; DB 18; Length 2256646;

Best Local Similarity 52.7%; Pred. No. 8.4e-07;
Matches 148; Conservative 0; Mismatches 130; Indels 3; Gaps 1;

Qy 10 CCAGTGTTCGATAATTATCCCTACCTTCAATGACGGGTGACGCTGCAAGCCTGCGCTC 69
Db 830640 CGTTTGGTCAGCATCATATTCCGGTGTAAGTTCGAGAAAGTTCTTCGACGAATGCGTG 830699
Qy 70 GGAACATCGTGGCGACACCTACCGGAAGTGAAGTGTCTTGTGACGCGGCTTCG 129
Db 830700 GCGTCCGTTGCTGCGCAGCGTACGCCAATCTGGAGATTCTGCTGTCGACGACGGCTCG 830759
Qy 130 ACCGATCGGACCTCGACATCGCAACAGTTTCCGCCCGGAACCTCGGCTCGCGACTGGTC 189
Db 830760 CCGCAATTTGCCGCGCATGTGCGACGCT---GGGCGCGCGGACCCGGCGATACGC 830816
Qy 190 GTTCAAGCGGCGCGATGATGGCCCTTACGACGCCATGAACCGCGGCTCGGCGTAGCC 249
Db 830817 GTGATTCAAGCCCAACAGCGGCGCTGTCCGAGCCCGCCAACTCCGATGCTCGTAGGCT 830876
Qy 250 ACAGCGGAATGGTACTTTTTTAGCGCGCGACGACACCT 290
Db 830877 ATCGCGCGGTACATATCTTCGCCGATTCGCGACGACCGT 830917

RESULT 13

US-10-193-002-306/c
; Sequence 306, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.

; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 306:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 306:
US-10-193-002-306

Query Match 6.9%; Score 55.6; DB 15; Length 1060;
Best Local Similarity 50.2%; Pred. No. 1.6e-07;
Matches 121; Conservative 22; Mismatches 95; Indels 3; Gaps 3;

Qy 338 ATCGCGCAAGCCATCTTGTCTATGCGGATGTTGATGCGTTCGACGAAAAGCGGCATG 397
Db 744 ANNMGACCCNAAKATTGNAWMMGGAAVNTWTMTGSGTTC-ACCAATTTCGGGGG 686
Qy 398 CCGGACCTTTCGACCTCGACCGCTCCTATTGAGACGAATTTGTGCCCAATCGATCT 457
Db 685 GGGNCCTTTGTCCTTTCGACCGCATCGGCCCTTACAACTCGGCTACCGAGTCTGGG 627
Qy 458 TTTACCGCGGTGAGCTTTTCGACCGCATCGGCCCTTACAACTCGGCTACCGAGTCTGGG 517
Db 626 TYACNCCCGGATTTTGGGCAMCATTCGTTCCCMCMCTCCGAAACNGGTTCTKTCG 567
Qy 518 CGGACTGGGACTTCAATATTGCTGCTTCTCCAAACCGCGCGCTGATTACCGCGCTACATGG 577
Db 566 CCGANTGGGANTTCAATATTGGNTGTTTTTCNAACCCARG-GTTGTTNACCSGNAACATGN 508
Qy 578 A 578
Db 507 A 507

RESULT 14

US-10-084-843-311/c
; Sequence 311, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.

;
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Twardzik, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 311:
SEQUENCE CHARACTERISTICS:
LENGTH: 1060 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 311:
US-10-084-843-311

Query Match 6.9%; Score 55.6; DB 15; Length 1060;
Best Local Similarity 50.2%; Pred. No. 1.6e-07;
Matches 121; Conservative 22; Mismatches 95; Indels 3; Gaps 3;

Qy 338 ATGCGGCAAGCCATCTGTCTATGCGGATGTTGTGATGCTTGCAGCAAGAAAGCCGGCATG 397

Db 744 ANNMGACCCNNAKATTTGNWVGGAAYNTRWTGSGTTC-ACCAATTNCSGGGG 686

Qy 398 CGGACCTTTTGAACCTCGACCGCCTCTATTGTGAGACGAATTTGGCCACCAATCGATCT 457

Db 685 GGCNCCTTTGCMCTTGGACCTTTTGTGTTNAGGSAATTS-KCCATNSNGGWWTT 627

Qy 458 TTTACCGCGGTGAGCTTTTCGACGCGATCGCCCTTACACCTGGCTACCGAGTCTGG 517

Db 626 TYACNCCCGGATTTTGGGCGAMCATTCGTTTCCCAACMCTTCGKACNGGTTCKTGC 567

Qy 518 CGACTGGGACTTCAATATTCGCTCTCTCCAAACCGCGGCTGATTACCGCTACATGG 577

Db 566 CCGANTGGGANTTCAATATTCGCTCTCTCCAAACCGCGGCTGATTACCGCTACATGG 577

Qy 578 A 578

Db 507 A 507

RESULT 15

US-10-282-122A-36601

; Sequence 36601, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA 034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 36601

; LENGTH: 1033

; TYPE: DNA

; ORGANISM: Salmonella paratyphi A

US-10-282-122A-36601

Query Match 6.9%; Score 55.4; DB 17; Length 1033;
Best Local Similarity 52.7%; Pred. No. 1.9e-07;
Matches 144; Conservative 0; Mismatches 126; Indels 3; Gaps 1;

Qy 22 ATATTATCCCTACTTCATATGACGCGTGACGCTGCAGCGCTGCCTCGGAAGCATCGTC 81

Db 28 ATCATTTGCCGTTTATAATATGCGGAGCGGATTTTAAATGCTTGCATGCGCTCGTTAATC 87

Qy 82 GGGCAGACCTACCGGGAAGTGGAAAGTGGTCTTGTGACGCGCGTTCGACCGATCGGACC 141

Db 88 GCGCAACGTTGTCGGCGCTGGAATATTATTGTGATGATGATGATGATGATGATGATGATG 147

Qy 142 CTCGACATCGCGAACAAGTTTCCGCCCGGAACTCGGCTCGCGACTGGTGTTCACAGCGGG 201

Db 148 GTTGAGATAGCAAAACATTACG---CGGAAATTACCGCATGTTTCGACTGCTTCATCAG 204

Qy 202 CCGCATGATGCCCTACGAGCGCCATGACCGCGCGCTCGCGCTAGCCACAGGCGAATGG 261

Db 205 GCCAATGCTGGCGCATCTGTGCGCCGTAATCTTGGCTTCAAGCGGCGACCGGCGATTAC 264

Qy 262 GTACTTTTTTTAGGCGCGACGACACCTCTTAC 294

Db 265 GTGCGCTTTGCGATGCGGATGACCTGCTTAC 297

Mon Mar 7 10:15:40 2005

us-10-805-311-23.rnpb

Page 10

Search completed: March 6, 2005, 23:25:32
Job time : 617 secs
